

The quest for relevance:
Research on compliance with EU law
/a very short version/

Dimiter Toshkov, Leiden University

Version: 23 October 2011

Introduction

It was not long ago when Ellen Mastebroek (2005) asked whether the state of research on European Union (EU) compliance is still a ‘black hole’¹. In fact, the publication of her review came in the midst of a true explosion in the number of articles and books devoted to the study of transposition, implementation, and compliance in the EU. Much light has been directed towards the ‘black hole’, but to what effect? It is the purpose of this article to review the empirical scholarship on EU compliance that has been published over the last ten years using two new research tools that allow systematic comparisons of existing studies – the Implementation² and Compliance³ online databases.

My first conclusion is substantive. Beyond the cacophony of approaches, methods and findings, a limited but coherent picture of EU (non)compliance emerges: At a very general level national administrative capacity and effective administrative co-ordination enhance formal and practical implementation, but EU rules can ignite opposition from various actors and give rise to conflict which often leads to delayed, poor, and incorrect application. Domestic institutions define the consent of which actors is necessary for compliance and in that sense they are important as well. On the other hand, the opposition of no single domestic actor (nor policy misfit) can be linked in a systematic way to noncompliance.

My second conclusion is methodological. The research design of quantitative and qualitative studies has not been optimized towards the discovery of the causal factors explaining compliance. Furthermore, very few of the factors on which the quantitative and qualitative literatures focus can be thought of as ‘interventions’ – variables that can be manipulated in the course of designing compliance mechanisms. As a result, despite the surge in empirical research on EU compliance, the contribution of the resulting scholarship to policy making and its influence to the broader social-scientific community has been limited.

The problems with deriving causal inferences from observational data are well-known (see for example King et al., 1994) but have not been properly addressed

¹ Mastebroek was echoing Weiler’s concern expressed back in 1991(Weiler, 1991).

² <http://www.eif.oeaw.ac.at/implementation/>

³ <http://www.eif.oeaw.ac.at/compliance/> Both databases have been developed with the help of the Institute for European Integration Research of the Austrian Academy of Sciences.

by the body of quantitative literature on EU compliance⁴. The direct regression approaches used to analyze the data are better suited for summarizing the observed patterns rather than supporting causal interpretations. Let me give one example that would clarify the point. The conclusion that lower levels of federalism are associated with lower levels of compliance seems rather well-established (see above). Can we interpret this association causally however? First, the variation in this institutional feature is rather limited in the European states being studied. Second, and more importantly, there is almost no variation over time in this variable, since these institutional features of the state are exceptionally stable. So the conclusion that federalism (or corporatism, national political culture, etc.) is associated with lower compliance is entirely based on the cross-sectional comparison between the average compliance in a couple of countries (Germany, Austria, etc.) and the average compliance in the remaining ones. It is obvious that Germany might have a different compliance record than the rest of the EU states for a myriad of reasons which might or might not have to do with federalism. The response of quantitative research to this ‘omitted variable’ problem is to include other covariates that adjust for the ‘other’ factors that influence Germany’s compliance performance⁵ but the success of this strategy crucially depends on the ability of the researcher to fully ‘control’ for all these potential confounding variables. In research designs that span long time periods and different policy sectors, we should be skeptical that this assumption is satisfied. So it would seem a good idea to focus on a very small set of very similar directives implemented in a very small set of very similar countries which only differ with respect to the (institutional) feature we are interested in. The problem of this strategy, however, is that we can not be sure whether the estimated causal effects that we find can be generalized beyond the specific circumstances of the test (the assumption of unit homogeneity). If compliance works according to a different logic in social policy than in transport policy (or in the 2000s vs. the late 1980s), we would be wrong to

⁴ The statistical literature on EU compliance has undeniably progressed over the last decade – multivariate models have replaced bivariate correlations, count models have replaced linear regression, survival models have been used to accommodate problems of censored data, semi-parametric survival models have replaced Weibull survival models to allow more flexible assumptions about baseline hazard of compliance, etc. The problems discussed in this article, however, relate to a more fundamental problem of deriving causal inference from observational data that cannot be addressed by adjusting the distributional form of the error terms or allowing for time-varying coefficients of the independent variables.

⁵ Including country dummies in the regression equation only makes the problem worse since it takes away the existing variation between, say, Germany and the remaining states, so the estimated effect is based on the extremely limited over-time variation.

transfer causal inferences from one policy field (or time period) to another. Quantitative researchers should be torn between the requirement for ‘isolating’ the causal effect on the one hand and the need for generalizable inferences on the other. The exclusive dominance of direct regression serves neither side of this double constraint well. Using a heterogeneous sample (Commission and ordinary directives, different time periods and policy sectors, etc) makes the estimated effect uncertain and potentially biased, while there is no indication how the estimated effect *varies* across these sub-samples.

There are several research strategies that can address the problem. The first is the use of matching (pre-processing that ensures that the ‘treatment’ and ‘control’ groups are truly comparable) (Ho et al., 2007) but this would come at the price of abandoning regressions that claim to test a bunch of causal hypotheses at once. The second is the use of multi-level models (Gelman and Hill, 2007) which explicitly model how the estimated effect changes within different samples (countries, etc.). None of these techniques can solve the problem of limited variation provided by the world with regard to some institutions. We might never know whether federalism has anything to do with the different transposition record of Germany, but matching and multi-level modeling can provide some leverage towards estimating the causal effects of the variables which allow for that.

It is worth discussing how case studies can contribute to the goal of identifying causal factors accounting for variation in compliance performance. In principle case studies can be highly valuable for explanatory research in shedding light on *causal mechanisms*, serving as *plausibility probes* for new theories or *critical tests* of well-established ones, and in providing more *valid measures* for a moderate number of cases vs. the often indirect proxies used by large-N studies (Adcock and Collier, 2001; Brady and Collier, 2004; Collier, 1995; Gerring, 2004, 2007a, b; Mahoney and Goertz, 2004). Overall, however, the case studies of compliance have not been very helpful to advance explanations of compliance. There are several good examples of case studies as plausibility probes used to illustrate the working of formal models . These plausibility probes, however, have not been followed by more systematic empirical research that directly tests the theoretical models⁶. Very few comparative

⁶ Many variables *indirectly* suggested by the models, like the policy specific veto players index, have been tested, but this is not a sufficient test for a model that put the *preference configurations* center-stage in explaining noncompliance.

case studies are designed based on a logic of case selection that allows to isolate causal effects. Parallel to the case of quantitative research, qualitative analysis should strive to approximate as close as possible the counterfactual situation (same context, different value of the main explanatory variable), which requires selection on the independent variable. Instead, cases are often selected on the basis of selection on the dependent variable (only compliance failures) which might be useful to generate new hypotheses but these studies are sold as theory-testing analyses which they cannot be. Also, comparative case studies often strive to cover different institutional contexts (e.g. federal and unitary countries) which is beside the point when one wants to *isolate* a causal effect. Furthermore, many comparative case studies fail to consider strong clues suggesting the importance of a certain variable if the direction of the effect is different than the one expected by the researcher (e.g. ‘misfit was high in all cases but compliance was timely, so misfit has no effect’ – any systematic analysis would need to conclude that misfit is positively associated with compliance but case studies seem to be able to get away with a conclusions of ‘no impact’). It is interesting that none of the single case studies in the literature explicitly positions itself as a critical theory test (either most-likely or least-likely). Most combine an intention to *test* a battery of existing hypotheses while simultaneously identifying *ad hoc* additional factors that influenced implementation and enforcement. While such approaches can be valuable in the early stages of a research program, they certainly cannot *test* existing arguments and consolidate the findings in a mature literature. Lastly, the recently conducted detailed review of case studies of EU compliance revealed that the purported benefits of case studies for illuminating causal mechanisms and improving on measures are seldom realized, mostly due to little explicit consideration of variables and hypotheses (an important exception in this regard is Falkner et al. 2005). In sum, the potential of case studies to contribute to the explanatory research on EU compliance (hence, research that is useful for policy makers) has not been fully realized so far, despite the multitude of case studies published. In my opinion, the main reason for this is the inattention to how the case studies fit into the existing body of knowledge already accumulated for the workings of EU implementation and compliance⁷.

⁷ Several contributions to the literature are based on mixed designs (Berglund, 2009; Kaeding, 2007; Luetgert and Dannwolf, 2009; Mastenbroek, 2007).

My last point is that even if scholars can increase the reliability and validity of their causal inferences about the determinants of compliance, the research will still remain of limited practical significance if the variables we study cannot be subject to intervention. Even if research convincingly shows that federalism *leads* to higher transposition delays, for example, there is little policy makers can do about it. Still that might be useful at least to zoom-in on the problematic part of the process. Many of the variables that we currently study, however, cannot be interpreted as intervention even in this very indirect way. For example, scholars can probably agree that Commission directives are transposed faster. But do we imply that if a directive adopted under co-decision were to be adopted by the Commission, its transposition would have been faster? I do not think so. It is rather the underlying properties of the issue that determine whether a directive is adopted by the Commission or under co-decision *and* whether it is transposed on time or not. So the inference that Commission directives are associated with shorter transposition, even if valid, is less useful than we would like to imagine. The same reasoning goes for variables like the number of national implementing measures, or the directive's length, etc. While it might be important for control for these variables, we should be careful in interpreting causally their associations with compliance.

Research on EU compliance better disentangle its two ambitions – to illuminate in a comprehensive way the state of transposition and implementation and to identify important causal factors of compliance failures. The first ambition might be served well by broad but heterogeneous samples and case selection by practical relevance and convenience. The second one is better addressed by more focused comparisons, closely matched samples and multi-level analyses which would provide less sweeping but ultimately more policy-relevant generalizations.

References

- Adcock R and Collier D (2001) 'Measurement Validity: A Shared Standard for Qualitative and Quantitative Research', *American Political Science Review* 95(3): 529-46.
- Berglund S (2009) *Putting Politics into Perspective. A Study of the Implementation of EU Public Utilities Directives*. Delft: Eburon.
- Brady H E and Collier D (eds) (2004) *Rethinking Social Inquiry. Diverse Tools, Shared Standards*. Oxford: Rowman and Littlefield.
- Collier D (1995) 'Translating Quantitative Methods for Qualitative Researchers: The Case of Selection Bias', *American Political Science Review* 89(2): 461-6.
- Gelman A and Hill J (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.
- Gerring J (2004) 'What Is a Case and What Is It Good For?', *American Political Science Review* 98(2): 341-54.
- Gerring J (2007a) *Case Study Research. Principles and Practices*. Cambridge: Cambridge University Press.
- Gerring J (2007b) 'Is There a (Viable) Crucial-Case Method?', *Comparative Political Studies* 40(3): 231-53.
- Ho D E, Imai K, King G and Stuart E A (2007) 'Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference', *Political Analysis* 15(3): 199-236.
- Kaeding M (2007) *Better Regulation in the European Union. Lost in Translation or Full Steam Ahead? The Transportation of EU Transport Directives across Member States*. Leiden: Leiden University Press.
- King G, Keohane R O and Verba S (1994) *Designing Social Inquiry. Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
- Luetgert B and Dannwolf T (2009) 'Mixing Methods: A Nested Analysis of EU Member State Transposition Patterns', *European Union Politics* 10(3): 307-34.
- Mahoney J and Goertz G (2004) 'The Possibility Principle: Choosing Negative Cases in Comparative Research', *American Political Science Review* 98(04): 653-69.
- Mastenbroek E (2005) 'EU Compliance: Still a 'Black Hole'?', *Journal of European Public Policy* 12(6): 1103-20.
- Mastenbroek E (2007) *The Politics of Compliance. Explaining Transposition of EC Directives in the Netherlands*. Wageningen: Ponsen & Looijen.
- Weiler J H H (1991) 'The Transformation of Europe', *Yale Law Journal*(100): 2403-83.