

TAKING STOCK: A REVIEW OF QUANTITATIVE STUDIES OF TRANSPOSITION AND IMPLEMENTATION OF EU LAW

Dimiter Toshkov¹

Paper to be presented at the NIG Annual Work Conference,
Maastricht, 25-26 November 2010

Abstract

This paper presents a literature review of all quantitative (statistical) studies of compliance with EU law. The paper introduces and makes use of a new online database <http://www.eif.oeaw.ac.at/implementation/> which presents a detailed and comprehensive overview and classification of the existing quantitative research on transposition and implementation of EU directives in the member states. The study discusses and compares the different conceptualizations and operationalizations of compliance used, the list and specifications of the explanatory variables included in the models, the hypotheses proposed, and, most importantly, the findings of the literature. While the academic field has made progress in terms of assessing the scale and dimensions of the transposition failures in the EU, the causal inferences advanced in the existing literature are often weakly supported and sometimes contradictory when all studies are considered. The literature review suggests that only causal relationships that are specific for a certain time period, policy area, country, or type of legislation can be supported by empirical data, which means that broad generalizations about compliance in the EU might be impossible to uncover. The paper also suggests that decomposing the implementation process into its component stages, incorporating more rigorously the interactions between the Commission and the member states, and paying closer attention to the multilevel structure of the data in the statistical models can benefit future research on compliance in the EU.

¹ Institute of Public Administration, Leiden University, e-mail: dtoshkov@fsw.leidenuniv.nl

Introduction

Over the last decade the academic study of the interactions between the European Union (EU) and the member states has flourished. The objective of this paper is to review the current state of the art of the quantitative (statistical) literature on compliance, transposition and implementation of EU law. The text is a companion paper to the online database of studies of implementation², available at <http://www.eif.oeaw.ac.at/implementation/>. The database presents a comprehensive and detailed overview of all published articles (and some working papers³) on the topic. The paper will necessarily focus only on a few aspects of the literature and will demonstrate the added value that the Implementation database offers for assessing the state-of-the-art in this academic subfield.

The study of compliance with EU law has attracted and continues to attract considerable attention. Quantitative research on the topic represents only a part of the existing work in a field where methodological pluralism thrives and the research designs and methods employed range from case studies (e.g. Falkner et al., 2005, Ben, 2009, Haverland, 2000, Steunenberg, 2007, Hartlapp, 2009), to qualitative comparative analysis (Dimitrova and Toshkov, 2009, Sedelmeier, 2009, Kaeding, 2008b), to nested analyses based on mixed methods (Toshkov, 2009, Luetgert and Dannwolf, 2009, Kaeding, 2007, Mastebroek, 2007). By focusing only on quantitative studies, this paper does not wish to imply that research based on other methods is unimportant (for recent reviews see Mastebroek, 2005, Treib, 2006). Due to the very methodological diversity and breadth of the field, however, an overview that attempts to cover all work will be monumental in span and ambition. A review focused exclusively on quantitative analyses has the advantage of having a well-defined scope. Moreover, the nature of statistical work allows for a more structured and detailed comparison of the various studies that covers the precise analytical techniques used, the list of variables included, conceptualizations and operationalizations, and results.

The sheer number of quantitative studies of compliance published over the last decade makes an up-to-date, comprehensive and regularly updated literature overview a worthy endeavor. Nowadays, scholars who take up the topic of compliance in the EU have either to spend a considerable amount of their research time to identify, review and compare all the studies, or, by sidelining some of the published work, to make uninformed choices in their

² The database has been developed and is managed by the author with the support of the Institute for European Integration Research of the Austrian Academy of Sciences.

³ The database includes working papers in addition to the published academic articles. By doing so I hope to avoid the bias resulting from significant results having a higher chance of being published in journals than negative findings and achieve a more valid picture of the literature.

research designs. Furthermore, the accumulation of knowledge in the field can be hampered if no easy access to all published work is available.

By presenting the important aspects of the quantitative studies of compliance side by side, the database makes reviewing the literature easier. For example, each individual study presents its hypotheses and reports the results in terms of its specific dependent variable (e.g. transposition time, delay, or number of infringements started in a year). The database standardizes this information by reporting hypotheses and results in terms of a single outcome – compliance⁴. In addition, it enables taking stock of the current state of the literature and a summary of what we have learned over the last decade about the topic of compliance with EU law. As a consequence, the database encourages replication of existing analyses using different methods, operationalizations, or datasets. Using the database, one can also identify areas where qualitative process-tracing work might be especially useful to settle conflicting findings.

The current paper does not attempt to present a meta-analysis (in the strict sense of the term) of the quantitative literature on compliance. Of course, a formal meta-analysis would have been an important contribution to the academic field. Unfortunately, the studies use too different operationalizations of outcome variables and too different standards of presenting the statistical results (often not reporting sufficient information) in order to make a formal meta-analysis possible.

The rest of the paper is structured as follows. The next section discusses the fundamental assumptions of doing statistical research on compliance. This is followed by an overview of the literature structured along several dimensions. First, different operationalizations of the dependent variable and their pros and cons are discussed. Second, the domain of the existing analyses is summarized and compared. Third, the methodological techniques used in the literature are reviewed. Fourth, the findings of the literature in regard to several groups of variables – institutions, capacity, preferences, directive features – are presented and assessed. The final section of the paper outlines the main conclusions of the paper and suggests some likely developments for the future.

Compliance with EU law and quantitative research

Sidelining issues of definition, I use both compliance and implementation in a general sense that covers formal transposition, practical application and enforcement of laws and policies

⁴ Furthermore, the coding of independent variables and baseline categories of binary and categorical variables has been standardized.

(for a detailed discussion see König and Luetgert, 2009, Falkner et al., 2005, Toshkov, 2009). Compliance with EU law as a topic of academic interest is one of those rare cases of an important real world problem that allows for structured empirical research due to the nature of the process that generates the outcomes. The practical reasons scholars have turned to studying compliance with EU law are obvious since non-implementation undermines the internal market and brings very tangible consequences for companies and citizens in Europe. What makes the topic even more seductive for social scientists is the fact that the process of implementation of EU law in the member states resembles a natural quasi-experiment where the input (EU legislation) is processed synchronously by a number of politico-administration systems. In addition, the implementation of EU directives allows for discretion in the exact timing, scope and manner of application. The resulting patterns of implementation outcomes present a window into the workings of the different national institutions and a chance to distill the causal effects of a multitude of factors that vary in the different countries. Finally, the study of compliance with EU law has burgeoned because of the wide availability of data, collected by European and national authorities, on the issue. These three basic rationales for research on compliance with EU law have to be taken into account when interpreting the design and findings of the literature, because certain choices that the researchers make (like using a complex index as a dependent variable, or studying only a few EU laws in many countries) make more sense if one is interested in the topic of compliance only as a setting to test certain causal hypotheses, rather than in the substantive dimensions and explanations of implementation problems.

Altogether, more than 30 separate statistical analyses of national adaptation to EU law have been published in academic journals over the last decade. None of these studies, however, ventures into the question as to why *statistical* analysis of compliance is appropriate. The assumptions that ground the statistical approach to studying transposition and implementation are not spelled out. It might seem obvious that the application of statistical methods fits and benefits the study of domestic adaptation to EU law, but in fact several different rationales can be proposed, each based on different assumptions. First of all, the application of statistics can be justified with the aim of making descriptive inferences only. Measuring compliance performance is never perfect, so statistics can address what is essentially a measurement problem. Collecting data on a large number of cases improves our knowledge of the tendencies and patterns of compliance because measurement errors cancel out in the aggregate. Similarly, statistical techniques can be used as a data reduction tool,

summarizing information in a compact way but still with no intention to support causal inferences. Different versions of factor analysis fall within this category but the use of regressions can also be conducted with descriptive or compact representation purposes only in mind.

Secondly, apart from correcting measurement problems and summarizing complex data, researchers can use statistics to make causal inferences. There can be at least two, rather different, rationales to employ statistics for advancing causal claims. First, compliance can be considered an essentially stochastic (random) process which necessitates the application of statistical analysis. If there are no deterministic reasons for transposition delays and implementation failures, case studies and even small-N comparative work are unlikely to discover causal effects because causes are only probabilistic and increase the likelihood of an event without being neither necessary nor sufficient for its occurrence. If there is no causal chain that necessarily lead from a set of conditions to an outcome, process tracing case studies will prove of little value to the scientific inquiry of compliance. But statistics can discover probabilistic causes that make an event more or less likely given a certain combination of conditions.

Alternatively, scholars can assume that compliance performance is in principle deterministic although in practice there are so many possible causal factors and interactions between them that only statistical analysis can test for the effects in a large pool of cases. In my opinion, most, if not all, of the published statistical analyses of compliance are based on this last rationale – compliance is in principle deterministic (reasons exist for each and every implementation failure) but the multitude of possible causes necessitates a statistical approach to explanation. It is a pity that none of the articles reviewed discusses these ‘deep’ methodological assumptions because they reveal the fundamental understanding of compliance as a social process that a researcher has in mind. Even if left implicit, the justification for the use of quantitative analysis carries with it certain implications that limit the research design choices and the interpretation of the findings.

The assumptions needed to make casual inferences from regression and related techniques are also left implicit in the bulk of the literature reviewed. It is useful to briefly review these assumptions. First, studies that aspire to estimate causal effects need to assume *unit homogeneity* (King et al., 1994, 90)⁵. In the context of implementation research, the assumption of unit homogeneity entails assuming that the observations used for the analysis

⁵ “Two units are homogeneous when the expected values of the dependent variables from each unit are the same when our explanatory variable takes on a particular value.”

are governed by the same data-generating process. For example, the transposition of Council and Commission (delegated) directives must be assumed to follow the same causal process if the sample includes both types of legislation. Similarly, when the analysis covers a longer time span, we need to assume that there are no fundamental changes in the way implementation works over time. Given the increasing attention to the problems of imperfect transposition and implementation of EU law since the 1990s this assumption might be problematic. Simply put, implementation of EU law during the 1970s might be a completely different ball game than implementation in the 2000s and researchers need to consider possible violations of the unit homogeneity assumption when they include different types of legislation and a long time span in the analysis.

The second assumption for estimating causal effects is *conditional* independence (King et al., 1994, 94)⁶. This assumption requires that there is (a) no selection bias, (b) no omitted variables bias, and (c) no endogeneity between the dependent and the independent variables. The quantitative literature on EU compliance is generally attentive to these requirements. Nevertheless, potential biases are sometimes overlooked. For example, the application of bivariate analyses is suspect to omitted variable problems given the highly correlated nature of many institutional and other state-related variables. A close reading of the literature can uncover examples of possible violations of the endogeneity assumption and subtle selection biases. I will return to these problems when discussing conceptualizations and operationalizations of the dependent variables and the methods of analysis later in the text.

This section highlighted the fact that compliance in the EU resembles a natural quasi-experiment allowing a window in the workings of different national politico-administrative systems as the main motivation behind research on EU transposition and implementation. It also concluded that statistical research is employed due to the great number of potential determinants of implementation performance and the complex nature of the interactions between these factors. The literature on implementation is more ambitious than simply summarizing data and instead attempts to uncover causal relationships. This section introduced the two assumptions needed to make causal inferences from observational data – unit homogeneity and conditional independence. The next section of the paper will move to the in-depth review of the various elements of the quantitative analyses – dependent variables, domain of analysis, analytical techniques, and findings. I will commence with the discussion

⁶ “Conditional independence is the assumption that values are assigned to explanatory variables independently of the values taken by the dependent variables”.

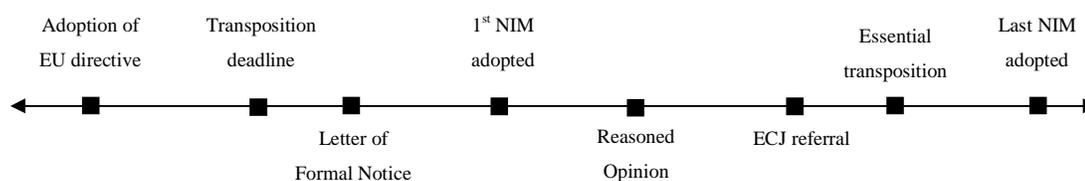
of the details of operationalization and measurement of compliance – the outcome analyzed in the reviewed literature.

Detecting compliance: operationalizing and measuring the dependent variable

Almost all of the studies reviewed employ two types of data with regards to the dependent variable: the transposition of EU directives in the member states, and infringement procedures against the member states for violations of EU law. The remaining analyses combine data from these two types into indexes of implementation performance or use reports by the Commission on implementation to operationalize compliance.

Figure 1 presents a timeline of the transposition of a hypothetical directive with different events indicated on the line. Most studies have focused on the adoption of the first national transposition measure as the relevant event defining compliance. Some analyses attempt to define when ‘essentially’ correct transposition has taken place, and few focus on the last transposition measure (that is the last one at the time of conducting the research) adopted. Alternatively, studies of infringements focus on the moment a Letter of Formal Notice, a Reasoned Opinion, or an ECJ referral, is sent with regards to the transposition and implementation of the directive.

Figure 1. The process of compliance with EU directives. *Note: NIM- national implementing measures.*



Both transposition and infringement data provide only a partial perspective on compliance. The main disadvantage of transposition data is that it only refers to the formal-legal part of the process. The main disadvantage of infringement data is that it is generated by strategic interactions between the Commission and the member states and is not a result of a process of perfect detection and pursuit of transgressions of EU law. Still, using infringement

procedures can sometimes improve estimates of compliance levels⁷. First, the start of an infringement procedure incorporates a qualitative assessment by the Commission on the adequacy of the notified national implementing acts in addition to the mere presence of any notified acts. Infringement procedures also reflect problems with the practical implementation of directives, so they give a glimpse beyond the formal-legal aspects of compliance.

Compliance is an irreducibly subjective (or ‘inter-subjective’) rather than an objective concept. Operationalizing and measuring compliance is not only imperfect because of measurement biases, insufficient information, etc. The judgment a researcher makes whether some member states have complied with a particular EU law is always open for criticism because there can be no objective standard of compliance that is applicable to all cases at all times. The fact that transposition and infringement data provide only partial perspectives on compliance has to be considered in this context. The shortcomings of transposition and infringement data should not be measured against some ‘perfectly’ objective measure because such a measure does not exist.

Following a strictly legalistic definition of compliance, as long as the European Court of Justice has not declared a certain national provision or practice in breach of EU law, we have to conclude that the provision or practice is compliant. Naturally, social scientists are not at ease with such a perspective because they know well that the process of detection, investigation and judgment of infringements is not infallible. At the same time, all interpretations of the compatibility or not of national provisions and practices with EU law cannot be separated from the actors who advance the interpretations. Non-governmental organizations, government departments, the Commission, advocacy coalitions, trade unions, etc. are all actors with a specific stake in the process which influences their perspective and interpretation of the state of compliance. Thus, qualitative data cannot provide a golden standard against which to judge the failings of the data on transposition and infringements that

⁷ Let us assume that we detect compliance either by (A) the absence of an infringement procedure, and (B) the presence of notified transposition measures. The indicator (A) will provide more valid estimates than indicator (B) if the Commission is more likely to (1) start an infringement procedure in case of notified measures but no real compliance than to (2) start an infringement procedure when there is compliance and notification. Situation (2) is not so implausible as it seems at first, if the member state complies, fails to notify right away, the Commission starts an infringement procedure, and only after that the member states notifies the transposition measure. The researcher looking back at the process will assess correctly the level of compliance by looking at the notified measures rather than by looking at the existence of an infringement procedure. Nevertheless, these cases should be few and far between since member states have an interest in reporting truthfully when they do comply. On the other hand, relying on infringement procedures can help us detect cases where a member state has not in fact complied but claims the contrary by submitting transposition measures (situation 1). These cases are quite probable given the incentive structure faced by member states.

quantitative work on compliance employs. We should, rather, inquire into the details and strategic setting of the processes that generate the data and attempt to counter any potential biases and threats to its validity. But there will always be room for subjectivity when deciding whether a certain national act or program implements sufficiently the EU directive or not. The concept of ‘essential’ compliance addresses precisely this tension but it also cannot escape the inherent subjectivity of compliance.

Looking at the studies that employ transposition data we quickly note that they disagree about the proper way to operationalize transposition timeliness. While some studies take the first national transpositions measure adopted, others take the last transposition measure to signify compliance. Yet others construct a categorical variable and distinguish between levels of transposition delay. Given the discussion in the previous paragraphs, all we can say about the appropriateness of the different operationalizations is that they depend on the purpose of the researchers and that they all carry certain benefits and problems. Relying on the first transposition measure often underestimates the problems with compliance while relying on the last one can overestimate the transposition delay.

In more technical terms, studies of transposition operationalize their dependent variable either as transposition time (the time between the adoption of the EU directive and some transposition measure), transposition delay (the time between the transposition deadline and the adoption of some transposition measure), transposition timeliness (a binary variable tracking whether transposition has been on time or not), or some categorical variable. Reviewing the literature, it seems that these choices have little impact on the conclusions of the study. Transposition time and delay are often highly correlated. Studies of the length of transposition delay that focus only on the delayed cases, however, suffer from selection bias. Unless the process that leads to some cases being delayed and others not is modeled, we should not accept at face value general conclusions about compliance reached by analyzing the length of delay based on data only on the cases that were delayed. Binary and categorical variables, by ignoring some of the information, are more conservative approaches that assume that the data is not reliable enough to allow us to treat transposition on a continuous scale, and that a delay of one week is substantively different from a delay of two years.

Practically all studies focus on the temporal aspect of compliance (but see Franchino and Hoyland, 2009)⁸. All analyses focus on explaining the *mean* of the distribution of the

⁸ Using the data on transposition from EURLEX, Franchino and Hoyland (2009) present an analysis of the involvement of national parliaments in the transposition process. This study is not included in the Implementation database, however, because the dependent variable is not compliance.

implementation outcomes. The *variance* of implementation performance – a variable that is interesting and important in its own right – is modeled only by Toshkov (2007a).

Moving from operationalization and measurement to data source issues, we note that most of the transposition studies use the CELEX (EURLEX) database. This data source has been extensively and rightly criticized as insufficient and unreliable (Hartlapp and Falkner, 2009). We have to recognize, however, that many of the actual analyses of transposition use CELEX (EURLEX) only as a first step in collecting the data and complement it with other national or EU-level databases. The main disadvantage of CELEX (EURLEX) is that it is essentially a database of transposition notifications and the researcher has to assume that the notifications present a close representation of the actual state of transposition. The database leaves the question whether an absence of notified measures signifies no need for notification (thus, full compliance), failure to notify the transposition measures, or a failure to adopt any transposition measures (thus complete non-compliance) (see König and Luetgert, 2009 for a detailed discussion).

The situation has improved in recent years with some standardization in the requirements for reporting national data to EURLEX, but the problem remains. No doubt, combining data from different sources increases its validity and reliability but the disclaimer that there is no final, objective interpretation of compliance should be kept in mind. For example, the relevance of some acts reported in the context of transposition of a certain directive is often open for discussion.

The very real limitations of using transposition data have led many researchers to focus on infringements instead. Usually, some aggregation of the number of infringement cases against a country over a period of time is taken as the dependent variable. Since the infringement procedures are composed of several stages, scholars have the choice to focus either on the initial stages (Letter of Formal Notice and Reasoned Opinion) or on the actual judgments of the ECJ. Because individual data on infringements is not easily available, the bulk of the research resorts to aggregation. Aggregated data, however, may lead to serious problems for deriving proper estimates of the relationships between variables in the statistical models (e.g. due to auto-correlation from one year to the next). In addition to these ‘technical’ concerns we have to remember that the Commission is an actor with limited capacity and with specific institutional interests in the infringement process and that by focusing on the cases it chooses to pursue we might get a distorted picture of compliance in the EU. The infringement procedure is a game between the national authorities and the Commission, and the data

generated by each move reflects strategic considerations and imperfect information. In addition, the data sources used to get information on the number of infringements resolved at different stages has its own shortcomings – the number of letters of formal notice is inconsistently reported which might bias conclusions about the rate of resolution of certain types of cases, etc.

In principle many of the issues raised above can be solved when individual-level data is available. However, we run into very serious problems with selection bias when we analyze individual-level infringement data if we focus only on the infringements. Data on which cases that led to a letter of formal notice also received an ECJ judgment is important in its own right, but it cannot be used to derive statements about compliance in general, because it only concerns cases that *have* been delayed or improperly implemented. Country rates of *closing* infringement cases are important but they cannot be taken as indications of country rates of compliance. A very simple example can demonstrate the point: the more cases you have at the early stages of the infringement procedure (e.g. because of delayed transposition), the more cases you will resolve before the ECJ judgment stage which will make you appear more compliant but only within the context of the infringement procedure and not in general, because you would still have more problematic cases at each stage of the procedure.

In conclusion, quantitative analyses of compliance have relied on two types of data and a wide number of specific operationalizations of transposition timeliness and delay and infringement numbers and occurrence. Since all these operationalizations provide only a partial look at compliance, great care should be exercised in framing and interpreting the inferences of each individual study. By recording in detail the precise operationalizations of the dependent variables employed, the Implementation database makes this task easier. In addition to the dependent variables, the domain of the analysis can influence our conclusions about compliance. The next section of the paper turns to this aspect of the literature.

Defining the domain of analysis and its influence on the results

Due to data collection limitations virtually no analysis covers all potential cases of compliance. The studies are distributed unevenly in terms of countries, policy sectors and time periods. Surprisingly, given its marginal place in the corpus of EU legislation, social policy receives the lion's share of academic attention (a total of seven articles in the database focus exclusively on the social policy field). Because there is increasing evidence that compliance varies systematically across policy sectors (for a recent statement see Haverland

et al., n.a.), focusing attention on only a few sectors and sidelining others can bias conclusions about compliance with EU law and policies in general.

To date, there is only one study that looks at all 27 member states of the EU (Steunenberg and Toshkov, 2009). Most of the research has concentrated on the EU-15 (for the time before the Eastern enlargement but after the accession Sweden, Finland, and Austria). The choice of which specific countries to study is often made with regards to practical and data availability concerns which leads to some countries being overrepresented in the literature. Systematic differences in country transposition and implementation performance are also well established. What we know about compliance in Europe might be seriously influenced by the fact that we have studied some countries (Germany, the Netherlands, and the UK) more intensely than others.

In aggregated data, the performance of some countries over time appears clustered but in specific datasets the clustering is rarely supported by formal analysis (Falkner et al., 2005, Thomson, 2007, Falkner, 2007, Thomson, 2009, Toshkov, 2007a, Falkner and Treib, 2008). Because cross-sector variation is at least as important as cross-country variation in compliance (see the discussion of multi-level models below) this fact should not come as a surprise. Also, grouping new versus old member states cannot be supported by the data (Steunenberg and Toshkov, 2009).

The final observation about the domain of analysis is that most of the studies focus on all directives within a policy sector or a country and do not filter important from unimportant EU legislation. Although very well established, this decision can be criticized on the basis of the considerable variability of issues covered by the same legal instrument – the directive. By analyzing all directives we pool laws on rear-view mirrors of tractors together with laws on anti-discrimination. Furthermore, Commission directives as such are more similar to government resolutions in domestic legal systems while regular (Council and co-decision) directives resemble primary legislation. Hence, putting Commission and regular directives in the same analysis might lead to problems with the assumption of unit homogeneity. We will have more to say about this issue later in the paper when we discuss the findings, but it is important to emphasize that the lack of strong causal inferences in the literature can be attributed possibly to the fact that many different types of legislation, adopted under the heading of a ‘directive’, are pooled into the datasets.

Analyzing the data: the statistical techniques

Moving from the domain of analysis to the discussion of the statistical methodology employed by the literature on compliance, two conclusions stand out: first, the statistical methods have become more sophisticated, and second survival analysis (and Cox proportional hazards) in particular is becoming the preferred way of analyzing transposition data.

Unlike other subfields of political science and public administration, the study of EU implementation has largely avoided some common problems with statistical analysis of political data. For example, when analyzing counts of infringements or delayed transposition cases, scholars have relied on the negative binomial distribution which is more appropriate for rare events (and over-dispersed data) than the normal distribution. Moreover, many of the recent analyses are well-reported, pay attention to substantive and not only to statistical significance, illustrate their results, and test some of the assumptions of the models. The use of interaction effects, which allow for the test of more complex and subtle hypotheses, is increasingly common as well.

Since the data on transposition is generated by observing a process over time, event history (survival) models are a natural choice for statistical analysis of the data. Initially parametric models were used (Mastenbroek, 2003), but more recently most scholars opt for the Cox proportional hazards version which makes fewer assumptions about the functional form of the baseline hazard of compliance.

Regardless of whether count models or event history analysis are preferred, explicit consideration of the multi-level structure of the compliance data is crucial for deriving valid inferences. As indicated in the previous section, implementation performance differs across countries *and* across sectors (and possibly over time as well). The country and sector dummies included more or less incidentally in different analyses provide enough evidence for this argument. However, it is difficult to find a single statement in the entire literature on the *amount* of variation in implementation performance that can be attributed to the different levels of observation – countries, sectors, and individual directives. Using the dataset analyzed by Haverland et al. (n.d.), we can find that approximately the same amount of variation is present at the country level and at the sector level.

Obviously, the transposition of two directives in the same policy field, and in the same country are not independent events as required by the assumptions of regression analysis. Many studies are aware of this problem and try to control for the multi-level structure of the data. Often, however, the fixes are only partial and address only one of the possible multi-

level complications. Moreover, the most often employed method of ‘fixing’ the problem – using panel corrected standard errors – is not enough to address the problem. The estimates of the effects, and not only the estimates of the standard errors, might be seriously off the mark if the observations are not truly independent, and in the presence of clustering and serial correlation. In addition, the relationships between independent variables and dependent variables can change depending on the context (country, sector, etc.). Including country and sector dummies in the regression models (the ‘default’ strategy to address these complications) allows a different starting level (intercept) for the relationship in different settings, but the relationship itself (the slope) is not allowed to vary. It is technically possible and theoretically justified to examine whether the relationships suggested by the literature are consistent among different countries and sectors, and not only in the aggregate (on multilevel models see for example Gelman and Hill, 2007).

Because of the increasing sophistication of the statistical techniques used to analyze compliance the findings also get more complex. While 10 years ago a statement like ‘EU approval is not related to transposition delay’ was typical, nowadays conclusions like ‘the effect of directive specialization on transposition time is negative during the first several weeks but gets positive and significant after this period’ are much more common. Interactions between explanatory variables, and interactions of the variables with time allow researchers to substantiate conclusions of increasing specificity which further leads us away from the search for broad generalizations about compliance in the EU (which might very well be futile).

Whatever the differences in techniques and model specifications, practically all reviewed studies advance causal claims based on macro-level regressions. Given the problems with making causal inferences from observational data, the use of regression to substantiate causal relationships (see for example McKim and Turner, 1997, King et al., 1994) is insufficiently reflected upon in the literature. The paper already mentioned that, in addition to unit homogeneity, researchers have to consider selection bias, omitted variable bias, and endogeneity concerns. Random selection and random assignment of ‘treatment’ conditions in an experimental setting can ensure that the estimates of causal effects are not plagued by omitted variables and endogeneity. Random assignment of directives to different ‘treatments’ (causal variables), however, is impossible. The researcher has to make use of the natural variation occurring in the real world. If we compare two samples that differ in terms of a ‘treatment’ (let’s say the novelty of a directive) we have to make sure that we control for all possible confounders (variables that are related both to the treatment and the outcome) in

order to make a valid inference about the effect of the treatment (Gelman and Hill, 2007). In practice, the list of possible confounders is long and usually not known in advance (e.g. the length, author, decision rule, salience, specialization, technicality, and scope of a directive are potential confounding variables with respect to the novelty of a directive and its impact of compliance). Within a regression context we address this problem by including all these potential variables in the model. We cannot ensure however that our two samples (treatment and no treatment, new and old directives) will be balanced even if we can assume that we have identified and measured all potential confounders. Maybe all our new directives happen to be overwhelmingly also Council directives adopted under unanimity while all old directives with very few exceptions are co-decision directives adopted under qualified-majority voting. Adjusting for the confounding influence of author and voting rule will be impossible, and we might not even notice the imbalance because the statistical software will still produce coefficients for all these variables (except in the rare cases of perfect colinearity). This situation is very likely when we include many country-level variables and must rely on only a small selection of countries (usually less than 15) in order to derive estimates of the effects of these country-level variables.

Matching is a procedure that can help alleviate these problems by reorganizing and restricting the original samples in order to achieve a better balance between treatment and control groups (Ho et al., 2007, Gelman and Hill, 2007). Matching ensures explicitly that each observation in the ‘treatment’ group is matched with an observation in the control group that is as similar as possible with regard to all possible important characteristics. Observations that do not fit are discarded. By using only the relevant information in the dataset, matching avoids some of the traps in making causal inferences from observational data but it has not been, so far, used in the literature on EU implementation. Different forms of matching have high potential for improving the methodological foundations of the field but they require that researchers refrain from macro-level regressions that provide estimates of dozens of ‘causal’ effects at a time, but rather focus on specific hypotheses and evaluate them with due care to the assumptions for making causal inferences.

Multi-level modeling and matching are two directions for methodological innovation in the field of compliance studies that can also help account for, and move beyond, the rather contradictory nature of findings. As the following section will make clear, few of the causal inferences suggested at one time or another in the literature have been consistently supported

and confirmed. More often, estimated relationships cover the full spectrum from positive to negative depending on the exact dependent variable, scope, and method used.

Taking stock: what affects compliance?

No less than 263 relationships between potential explanatory factors and some aspect of compliance have been tested in the articles reviewed in this paper and the accompanying database. In order to ease comparison of this enormous number of findings, the database on which this paper is based classifies the independent variables according to their level (EU, national) and the broad type or category that the variables falls into (e.g. institution, culture, etc.). While fallible, the process of categorization allows for taking stock of the findings of the literature. In this part of the paper, I will go through some of the types of variables recorded in the database. I will summarize the findings by indicating how many of the analyses report a significant negative (positive), a non-significant negative (positive) and no relationship⁹, separately for studies which use transposition, implementation, and combined index data. Where possible, I will suggest scope restrictions or domain limitations that can makes sense of the contradictions.

Institutions

Many institutional features of the European states have been probed as possible explanations of compliance. An incomplete list includes federalism, regionalism, corporatism, meso-level institutions like co-ordination strength of the executive, extent of parliamentary scrutiny on EU affairs, etc.

The impact of federalism/regionalism on compliance seems rather well-established. We can say quite confidently that the impact is not positive. Approximately half of the studies report a significantly negative relationship (Haverland and Romeijn, 2007, König and Luetgert, 2009, Thomson, 2007, Linos, 2007), the remaining ones agree that there is a negative relationship but cannot find statistical significance (Steunenberg and Toshkov, 2009, Giuliani, 2003, Mbaye, 2001, Jensen, 2007).

⁹ Some readers might be disturbed by the fact that non-significant relationships are not reported as ‘no relationships’. However, statistical significance and substantive significance are not the same. Especially in small samples, the lack of statistical significance might obscure a substantively important relationship. That is why only effects that are zero for all practical purposes have been reported in the ‘no effect category’.

The effect of corporatism has received much attention but the findings are inconclusive. A positive effect is discovered by König and Luetgert (2009) and a positive but not significant one is found by Kaeding (2006) and Thomson (2007), while Lampinen and Uusikyla (1998) and Mbaye (2001) find a non-significant negative relationship.

Table 1. Federalism and regionalism

	Negative (significant)	Negative	~ Zero	Positive	Positive (significant)
Transposition	5	3	0	0	0
Infringements	0	1	1	0	0

Table 2. Corporatism

	Negative (significant)	Negative	~ Zero	Positive	Positive (significant)
Transposition	0	1	0	2	1
Infringements	0	1	0	0	0

Table 3. Parliamentary scrutiny

	Negative (significant)	Negative	~ Zero	Positive	Positive (significant)
Transposition	0	0	0	1	2
Index	0	0	0	0	2

The extent of parliamentary scrutiny is positive and significantly related to compliance. Evidence for this link is brought by Bergman (2000), Giuliani (2004) and Linos (2007). Linos in fact finds a significant effect on the length of transposition delay but not on the occurrence of delay (transposition timeliness). Parliamentary involvement (measured not as an institutional characteristic but as the share of primary legislation in the transposition acts) is beneficial also according to König and Luetgert (2009).

Veto players

The concept of veto players plays a prominent place in research on Europeanization, and compliance with EU law more specifically. Although related to federalism, it is more dynamic than an institution as, depending on the precise operationalization, its values change with the number of parties in government, type of Parliamentary procedure, etc. In fact, it combines institutional and preference information in order to provide an index of the capability and capacity of the politico-administrative system to change laws and policies.

The literature strongly suggests that the impact of veto players is not positive, and most likely it is negative. Significant relationships are reported by Kaeding (2006, 2008a) Linos (2007) and Giuliani (2003). Weaker but still non-positive relationships are found by Jensen (2007), Toshkov (2007b), Mbaye (2001) and Kaeding (2006) with a different operationalization. Steunenberg and Kaeding (2009) qualify these findings by reporting an effect that starts positive but turns negative with the passage of time.

The closely related index of political constraints has negative and significant impact according to Perkins and Neumayer (2007), no impact according to Börzel et al. (2007) and positive and significant impact according to Hille and Knill (2006) who work with data on the candidate member states only. Government type (minority or not) has no effect (Bergman 2000), but Giuliani (2003) insists that the effective number of parties has a positive and significant effect. The number of parties in government has a negative and significant effect according to Toshkov (2007a, 2008).

Some of the less often tested institutional hypotheses concern policy centralization (no effect – Siegel 2006), executive control of the legislature (positive effect according to Giuliani 2003 but negative according to Siegel 2006), EU co-ordination strength (negative effect – Giuliani, 2004), oversight type (Jensen 2007), and consensualism (negative effect– Giuliani 2003).

Capacity

Next to institutions, variables capturing in one way or another government capacity and quality are analyzed by many of the works reviewed. There is rather strong evidence that government quality influences positively compliance (Berglund et al., 2006, Börzel et al., 2007, Haverland and Romeijn, 2007, Hille and Knill, 2006, Lampinen and Uusikyla, 1998, Linos, 2007, Mbaye, 2001, Toshkov, 2007a, Toshkov, 2008, Siegel, 2006). The few aberrations are Thomson (2007), Perkins and Neumayer (2007), and Steunenberg and

Toshkov (2009) (the latter work with a special sample of legislation and include the new member states in the analysis).

The picture is confirmed by studies who track the influence of levels of corruption on compliance and find a negative effect (Linos, 2007, Kaeding, 2006). The opposite conclusion by Mbaye (2001) might be driven by colinearity problems.

Table 4. Veto players

	Negative (significant)	Negative	~ Zero	Positive	Positive (significant)
Transposition	5	1	1	0	0
Infringements	0	1	1	0	0
Index	1	0	0	0	0

Table 5. Political constraints

	Negative (significant)	Negative	~ Zero	Positive	Positive (significant)
Infringements	1	0	1	0	0
Index	0	0	0	0	1

Table 6. Administrative efficiency

	Negative (significant)	Negative	~ Zero	Positive	Positive (significant)
Transposition	0	1	1	0	8
Infringements	1	0	1	0	3
Index	0	0	0	0	1

Table 7. Corruption index

	Negative (significant)	Negative	~ Zero	Positive	Positive (significant)
Transposition	3	0	0	0	0
Infringements	0	0	0	1	0

Different aspects of co-ordination capacity have been tested with the general conclusion that they have a significantly positive effect on compliance while the involvement of more than one ministry influences negatively transposition time (Mastenbroek, 2003, Haverland and Romeijn, 2007, Steunenberg and Toshkov, 2009) (see also the discussion of institutions).

More distant proxies of state power and capacity exhibit no effects. GDP and GDP per capita (Börzel et al., 2007), state financial capabilities (Hille and Knill, 2006) and fiscal resources (Mbaye, 2001) are not associated with compliance. Comparative economic power is negatively related (Mbaye, 2001). All in all, bureaucratic quality improves compliance but more general measures of state capabilities show no effect.

Preferences

The third major group of explanations refers to preferences and attitudes. There are numerous hypotheses advanced in the literature that try to capture the intuition that member states which have a stronger preferences the EU policies to be ‘downloaded’ will comply faster than states that oppose them. Many different proxies of state preferences have been tested: some of them refer to the attitudes of the public at large, other to government preferences. Some studies employ general EU or ideological preferences while others try to measure preferences with regard to individual policy sectors, or pieces of legislation. Altogether the findings are rather inconclusive.

Let us start first with *state power* – a factor that is not directly related to preferences but which is a proxy for the ability of member states to shape the EU legislation to their liking. The findings with regards to the effect of state power are rather polarized. Mbaye (2001), Jensen (2007) and Perkins and Neumayer (2007) find evidence for a positive and significant effect. On the other hand, Giuliani (2003), Siegel (2006) and Börzel et al. (2007) find a significant negative effect. The conflicting findings are even harder to explain given the relatively straight-forward operationalization of state power as some function of the number of votes in the Council of Ministers. Using an alternative measure – logged population - Perkins and Neumayer (2007) find no effect.

Moving from state power to societal EU attitudes, the conclusion is more straightforward – a more EU supportive public does not lead to a more compliant government. Significant negative effects are uncovered by Mbaye (2001) and Bergman (2000) while most of the studies find no significant relationship - Siegel (2006), Kaeding (2006), Börzel et al. (2007) and Lampinen and Uusikyla (1998). The only study that finds a significant positive

effect is based on data on the candidate countries from Central and Eastern Europe (Toshkov 2009). Hence, the conclusion that a society which harbors greater support for the EU does not lead to faster transposition and fewer infringements can be qualified: during Enlargement periods the candidate countries are sensitive to societal EU support and adapt faster to the EU requirements when support is higher.

Table 8. Bargaining power

	Negative (significant)	Negative	~ Zero	Positive	Positive (significant)
Infringements	2	0	1	0	3
Index	1	0	0	0	0

Table 9. Societal EU attitude

	Negative (significant)	Negative	~ Zero	Positive	Positive (significant)
Transposition	0	1	1	0	1
Infringements	1	0	2	0	0
Index	1	0	0	0	0

Table 10. Government EU position

	Negative (significant)	Negative	~ Zero	Positive	Positive (significant)
Transposition	0	1	0	1	2
Infringements	0	0	0	1	0
Index	0	0	0	1	0

Table 11. Government Left/Right position

	Negative (significant)	Negative	~ Zero	Positive	Positive (significant)
Transposition	0	3	0	1	1
Infringements	0	0	1	1	0

In addition to societal attitudes, scholars have also tested the impact of government ideological positions on compliance. While government affinity to the EU (measured usually with the help of expert surveys) seems positively related to compliance, left-right positions do not matter. All studies that do find a significant relationship between governmental EU positions and compliance are based on data from the period of Enlargement. Similarly, the only significant effect of left right ideological positions is again found in the context of EU accession (Toshkov, 2008). The remaining articles which analyze the relationship between socio-economic left/right positions and compliance do not report a significant effect (Jensen, 2007, Toshkov, 2007a, 2007b, Linos, 2007, Siegel, 2006).

More specific indicators of government preferences also demonstrate limited explanatory capacity. For example, Toshkov (2007b) finds a significant positive effect of government positions on EU social policy on the adaptation success of the candidate countries from CEE in the field of social policy. But using directive-level data Thomson (2007) concludes that there is no relationship between a country’s disagreement with a directive and compliance. With the help of expert-survey derived preference data on individual directives, Thomson et al. (2007) find a significant positive relationship between a country’s incentive to deviate and compliance with infringement procedures but a negative relationship with transposition delay. Zhelyazkova and Torenvlied (2009) report no effect for the same independent variable and transposition delay. Linos (2007) reports a negative, but non-significant effect of a country voting against a directive on the length of transposition delay and on transposition timeliness for the social policy sector.

Table 12. Disagreement with the directive

	Negative (significant)	Negative	~ Zero	Positive	Positive (significant)
Transposition	1	2	2	0	0
Infringements	0	0	0	0	1

Table 13. EU-level conflict

	Negative (significant)	Negative	~ Zero	Positive	Positive (significant)
Transposition	2	0	0	0	2* (time dep)

Table 14. Misfit

	Negative (significant)	Negative	~ Zero	Positive	Positive (significant)
Transposition	4	0	0	2	1
Infringements	1	0	0	0	0

Next to country and directive level preferences and positions, some scholars use indirect proxies to capture the incentives a country has to comply with EU law. Most articles find no significant relationship, however. König and Luetgert (2009) report a non-significant negative effect of the domestic value-added share of the sector. Perkins and Neumayer (2007) find a positive and significant effect of intra-EU trade, but Siegel (2006) concludes that the same effect is negative and not significant. Unemployment levels are not related to compliance in the social policy sector (Linos 2007, Toshkov 2007b), neither are labour costs (Linos 2007). Surprisingly, net transfers from the EU have a negative and significant effect on compliance according to Perkins and Neumayer (2007) and König and Luetgert (2009).

Conflict: EU level and national level

Another way to account for the variation in transposition performance that is related to preferences is to examine the effect of conflict. Conflict at the EU level when adopting the directive and conflict at the national level when transposing the directive have been hypothesized to exert an influence on the implementation patterns. With regards to EU level conflict, Zhelyazkova and Torenvlied (2009) find a complex effect which starts at zero but turns positive with the passage of time (using heterogeneity and policy polarization as measures). König and Luetgert (2009) and Luetgert and Dannwolf (2009) however argue for a strongly negative and significant effect of the size of the EU core

The exploration of domestic-level conflict brings us back to the discussion of veto players because essentially both institutional and preference-based arguments try to capture the idea that the more disagreement (and the more institutionalized opportunities for disagreement) you have at the national level, the more cumbersome compliance will be. The findings about the impact of domestic preference heterogeneity are relatively strong. König and Luetgert (2009) and Luetgert and Dannwolf (2009) discover a significantly negative relationship. Toshkov (2008) reports no effects for the ideological distance between

governing parties on both the EU and the left right dimensions, but a negative and significant effect for the number of parties in government (which is a more indirect proxy for conflict).

Misfit

The hypothesis that countries will deter compliance when they do not like the legislation they are about to implement (or ‘download’) is closely related to the expectation that when that the more the EU and the national policy differ, the more difficult the adaptation to the EU rules. This so-called misfit hypothesis is a central plank of studies of Europeanization and occupies a central place in quantitative studies of implementation with EU directives as well. The findings are relatively straightforward. Steunenberg and Toshkov (2009), Mastebroek (2003), Thomson et al. (2007) and Thomson (2007) report a significant negative relationship while Linos (2007) and Kaeding (2006) find positive although not significant effect (the exception is one of the two analyses that Linos presents in which the effect is positive and significant).

Directive-level features

Most of the variables discussed so far capture differences between states and within states over time. A significant amount of effort has been directed into analyzing the impact of certain features of a directive on its chances for timely and proper implementation. Although causal interpretations are often endorsed for these effects, it is actually quite hard to see how the differences in compliance likelihood between different types of directives can be interpreted as causal. For example, even if we find that Commission directives have a higher chance of being transposed on time, do we believe that if the same subject matter was to be adopted by the Council rather than by the Commission, compliance would have been different? It seems very hard to disentangle the reasons that make one directive more likely to be adopted by the Commission rather than by the Council (or new vs. amending, or adopted under unanimity vs. qualified majority voting) from the reasons that affect compliance with the directive. Especially using multivariate regression where the researchers hopes to control for all these different reasons by including many possible confounders in the model, interpreting differences as causal effects seems problematic.

The findings from the existing literature about the relationships between features of the directives and compliance, be they causal or not, are inconclusive. Beginning with *discretion* – the amount of leeway contained in the text of a directive – and its impact on

compliance we can immediately see that scholars disagree. While Steunenberg and Toshkov (2009), Kaeding (2008a) and Zhelyazkova and Torenvlied (2009) report a significant negative effect, Thomson et al. (2007) and Thomson (2007) find a positive one. It should be noted that on a theoretical level, the hypothesized impact of discretion also oscillates from negative to positive.

Table 15. Discretion

	Negative (significant)	Negative	~ Zero	Positive	Positive (significant)
Transposition	3	0	0	1	1
Infringements	0	0	0	1	0

Table 16. Voting rule

	Negative (significant)	Negative	~ Zero	Positive	Positive (significant)
Transposition	1	0	1	1	1
Infringements	0	1	0	0	0

Different concepts like complexity and salience have been operationalized and tested using the number of recitals of a directive as a measure. Kaeding (2006) finds a positive effect in the transport sector while Haverland and Romeijn (2007) and Toshkov (2008) report no significant effect for the social policy sector and the new member states respectively. The variable *political sensitivity* which is a factor positively related to the number of recitals, amongst others, is negatively related to transposition success according to Steunenberg and Kaeding (2009).

The novelty of directives (new vs. amending) appears to be related equally inconsistently with compliance. Luetgert and Dannwolf (2009) and Kaeding (2006) discover significant negative effects, while Haverland and Romeijn (2007) and Mastebroek (2003) find no significant effects at all. The voting rule according to which a directive has been adopted has a very inconsistent impact as well, according to the literature as can be seen from Table 16.

The studies seem to agree that Commission directives (delegated legislation) are more likely to be complied with (Luetgert and Dannwolf, 2009, Mastenbroek, 2003), or at least they are not less likely (König and Luetgert, 2009, Kaeding, 2006), apart from transposition in the context of Enlargement negotiations when the effect is negative and significant (Toshkov 2008). Furthermore, Steunenbergh and Kaeding (2009) find no effect for their *specialization* variable which is a factor score positively related to the number of annexes and shorter transposition deadlines of directives.

Summary

Summarizing the review of the findings so far, we can make a crude classification of the variables for which evidence for a relationship with compliance is strong and relatively univocal, those for which the findings are not very consistent but the exceptions can be accounted for relatively easily, and finally those variables for which the literature disagrees whether they do have an effect, and what the direction of the influence is.

The variables that we are almost certain affect compliance positively (or at least not negatively) are:

- (1) administrative efficiency,
- (2) parliamentary scrutiny, and
- (3) coordination strength.

The variables that exert a negative (or at least a non-positive) influence are:

- (4) federalism/regionalism,
- (5) corruption levels,
- (6) veto players,
- (7) number of ministries involved, and
- (8) domestic conflict.

We can be quite certain that (9) general economic indicators are not related to compliance as are (10) very general (economic) proxies for country preferences.

Societal (1) and (2) governmental EU attitudes and to some extent (3) governmental left/right positions seem to be positively related to compliance only during Enlargement periods but either have the opposite effect (possibly societal EU support) or no effect (government EU support and left/right positions) otherwise. The generally positive effect of (4) Commission directives on compliance is also reversed for candidate countries. We can

also add (5) *misfit* to this group as it seems to work in the general case but not in the transport and social policy fields

The evidence for the following variables is so mixed that we can hardly conclude anything about their impact:

- (1) corporatism,
- (2) political constraints,
- (3) type of government and number of parties in government,
- (4) bargaining power,
- (5) country disagreement with a directive,
- (6) EU level conflict,
- (7) discretion, and
- (8) directive voting rule.

This overview zooms in only on variables that have been analyzed by several studies. Of course, relationships that have been examined only once appear to have more consistent effects (this is not to say that these effects cannot prove to be strong and replicable).

Possibilities to extend the analysis

Many other relationships have been tested in the literature. Different aspects of national political culture, features of the national transposition measures, learning, etc. have all been analyzed at one time or another. The review of the findings of the quantitative literature on compliance can be extended further. The database keeps records on many other types of variables. The full list of categories is available on the website of the database. The database contains additionally information about the size of the estimated effects (where available) and details on the operationalizations of the variables so as to enhance replicability of the analyses. The review presented above is sufficient to demonstrate, however, the opportunities that the database offers for summarizing, assessing, and discussing the state of the art of statistical research on the transposition and implementation of EU law.

The review also makes clear that very few of the findings are corroborated by all studies in all settings (sectors, countries, and time periods). In fact, some of the findings that receive strong support are that directive features matter for compliance. Therefore restricting the analysis to only a subset of all EU directives and using even more interaction effects (and higher order interaction effects) can address the heterogeneity of the population (the EU directives) and account for the puzzle of conflicting findings. The use of interaction effects is

gaining prominence in the literature (e.g. Steunenberg and Kaeding, 2009, Zhelyazkova and Torenlvied, 2009, Luetgert and Dannwolf, 2009, Toshkov, 2008, Thomson et al., 2007) but can receive even wider application.

Conclusions: Why so much inconsistency?

Many of the findings reviewed above are inconsistent and not supported by different analyses. Given that the studies focus on different aspects of compliance (transposition, implementation), have varying domains (in terms of countries, sectors, and time periods), use different statistical techniques and model specifications, and employ different sets of independent variables (operationalized and measured differently) the disagreements in the literature are not that surprising. One way to deal with the inconsistency is to accept that our findings will be very different depending on whether, for example, we use the first or the last transposition measure reported, or we count letters of formal notice instead of reasoned opinions, etc. More appropriate operationalizations and more reliable data can certainly help in this regard. It should be noted, however, that some research design problems, like the choice of a dependent variable, cannot be objectively 'solved' and will always have a degree of subjectivity.

The fact that the estimated effects are so sensitive to these research design decisions is another argument in favor of changing the strategy from macro-level regressions estimating numerous effects simultaneously to matching procedures focusing on distilling the effect of one, or a small number of variables, at a time. Many of the country-level variables analyzed by the literature have very limited (if at all) variation over time and the number of cross-sectional units is also small (up to 27, but most often less than 15). Estimating simultaneously a number of country-level variables in this setting leads to a situation in which the statistical model has to extrapolate the data a lot in order to create the relevant 'counterfactual' for estimating the effects. For example, if most of the countries in our sample have a high number of veto players *and* negative public EU attitude while the others have a low number of veto players *and* positive EU attitude, and if we try to get estimates for the effect of both veto players and EU attitude variables, the numbers that the statistical software will produce would not mean much. Matching can improve on the situation by selecting only observations that bring relevant data.

In addition, we might have to accept that our findings have a very confined domain of application and conclusions reached on the basis of analyzing social policy cannot be

extended to other policy areas, for example. The plea for more attention to the multi-level structure of the data, expressed earlier in the paper, will recognize these limitations and might help to make sense of the data better than existing approaches. It might very well be that the relationships between different institutional, preference-based, cultural, enforcement, etc. factors and compliance on the other hand are real and waiting to be discovered, but that their complexity and ever-changing nature makes the task very difficult. The trend towards using more interaction effects can complement full-fledged multi-level modeling to address these challenges.

Next, we should be able to recognize the possibility that there might be just little to discover and that compliance in the EU might not reveal many systematic relationships. Similarly, even if the relationships are present, they might be so weak that the ‘signal to noise’ ratio prohibits us from estimating effects with any reasonable precision. The success of the literature to identify at least several relatively consistent relationships and the persistence of systematic variance between (some) countries and across (some) sectors indicates otherwise, but the possibility that there is not much to discover is always present.

Finally, in order to identify important relationships we need larger datasets and for that we need many observations covering extended periods of time. We have to assume, however, that the structure of the compliance process remain the same over time. In fact, there are many reasons to believe that the process of compliance changes drastically. It is also quite likely that all the attention of the academic community to compliance in the EU itself contributes to the transformation of the process. In fact, the 33rd study of compliance will analyze a process that has fundamentally different institutional setting, salience, actors involved, etc. than the first one. The object of research ‘reacts’ to the ‘scientific’ attention and interventions.

As the literature review has made apparent, the academic study of implementation in the EU based on quantitative methods is a growing and lively field that has offered a number of important conclusions. First, it has uncovered several robust associations between compliance and administrative efficiency and types of directives, for example. Somewhat accidentally, it has also demonstrated that many variables have only complex (time-dependent), contingent (dependent on the values of other variables) and domain-restricted (valid only for certain countries, sectors, or time periods) effects. There is visible progress in the methodological sophistication and increasing reporting standards of the literature. Several challenges, however, remain.

First of all, the combined results from all studies indicate that maybe a theory and a model of compliance with *all* EU directives is not possible. Simply, the body of EU legislation is too heterogeneous to allow such a single theory and statistical model. How and when to filter for certain aspects of the directives is an open question, but it is likely that there will be no single explanation that can account simultaneously for the implementation of legislation on the size of rear-view tractor mirrors and for the implementation of legislation on the freedom to provide services in the internal market.

Second, the multi-level structure of the data has not been modeled explicitly and comprehensively. We could be missing important effects because they have different direction in different settings (e.g. countries) or we could be generalizing some relationship that holds strongly only in one sector for example, to the entire body of EU law. Even a simple statement about the amount of variation that can be attributed to the different levels of the data is missing from the literature at the moment.

Finally, methodological sophistication can only lead us a certain way towards progress in this academic field. The key for better empirical findings lies with better theorizing of compliance in the EU. Linking the institutional, preference, and capacity-related domestic variables with the strategic interactions between the Commission and the member states in the context of the EU enforcement procedure is a pressing need. Zooming in at an even greater level at the *process* of national adaption to the EU requirements provides another avenue for progress. Many of the hypotheses suggested by the literature can be more fruitfully tested if more detailed data on the process of implementation is available. For example, for now hypotheses about the effect of executive co-ordination are tested on data recording the time until the final adoption of national transposition measures (which often includes time spent for parliamentary discussions, presidential vetoes, infringement challenges, etc.). If data on the time until the drafting of the bill at the government were available, that would allow a much better test of the effect of executive co-ordination. Similarly, hypotheses about the impact of the veto players cannot be efficiently tested on data that includes many cases which do not even require the agreement of these ‘veto’ players.

In conclusion, using the Implementation database, which is going to be regularly updated, the field of EU compliance studies can better assess its contributions and shortcomings. Future research on EU compliance and implementation will have the benefit of easy access to and comparison of existing studies, a quick reference to operationalization details, and, hopefully, a source of inspiration.

References

- Ben, C. (2009) The Second Time as Farce? The EU Takeover Directive, the Clash of Capitalisms and the Hamstrung Harmonization of European (and French) Corporate Governance*. *JCMS: Journal of Common Market Studies*, 47, 55-79.
- Berglund, S., Gange, I. & Waarden, F. v. (2006) Mass Production of Law. Routinization in Transposition of European Directives. A Sociological-Institutionalist Account. *Journal of European Public Policy*, 13, 692-716.
- Bergman, T. (2000) The European Union As the Next Step of Delegation and Accountability. *European Journal of Political Research*, 37, 415-429.
- Börzel, T. A., Dudziak, M., Hofmann, T., Panke, D. & Sprungk, C. (2007) Recalcitrance, Inefficiency and Support for European Integration. Why Member States Do (Not) Comply with European Law. *CES Working Paper, Harvard University*
- Dimitrova, A. & Toshkov, D. (2009) Post-accession compliance between administrative coordination and political bargaining. *European Integration Online Papers (EIoP)*, 13.
- Falkner, G. (2007) Time to Discuss: Data to Crunch or Problems to Solve? A Rejoinder to Robert Thomson. *West European Politics*, 30, 1009 - 1021.
- Falkner, G. & Treib, O. (2008) Three Worlds of Compliance or Four? The EU-15 Compared to New Member States. *JCMS: Journal of Common Market Studies*, 46, 293-313.
- Falkner, G., Treib, O., Hartlapp, M. & Leiber, S. (2005) *Complying with Europe? The Impact of EU Minimum Harmonisation and Soft Law in the Member States*, Cambridge, Cambridge University Press.
- Franchino, F. & Hoyland, B. (2009) Legislative Involvement in Parliamentary Systems: Opportunities, Conflict, and Institutional Constraints. *American Political Science Review*, 103, 607-621.
- Gelman, A. & Hill, J. (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge, Cambridge University Press.
- Giuliani, M. (2003) Europeanization in Comparative Perspective: Institutional Fit and National Adaptation. IN Featherstone, K. & Radaelli, C. M. (Eds.) *The Politics of Europeanization*. Oxford, Oxford University Press.
- Giuliani, M. (2004) UE Compliance: Macro, Meso or No Institutions At All? *URGE Working Paper*.
- Hartlapp, M. (2009) Implementation of EU Social Policy Directives in Belgium: What Matters in Domestic Politics? *Journal of European Integration*, 31, 467 - 488.
- Hartlapp, M. & Falkner, G. (2009) Problems of Operationalization and Data in EU Compliance Research. *European Union Politics*, 10, 281-304.
- Haverland, M. (2000) National Adaptation to European Integration: The Importance of Institutional Veto Points. *Journal of Public Policy*, 20, 83-103.
- Haverland, M. & Romeijn, M. (2007) Do Member States Make European Policies Work? Analysing the EU Transposition Deficit. *Public Administration*, 85, 757-78.
- Haverland, M., Steunenberg, B. & van Waarden, F. (n.a.) Sectors at different speeds: Analyzing transposition deficits in the European Union. *Journal of Common Market Studies*, forthcoming, n.a.
- Hille, P. & Knill, C. (2006) 'It's the Bureaucracy, Stupid': The Implementation of the Acquis Communautaire in EU Candidate Countries, 1999-2003 *European Union Politics*, 7, 531-552.
- Ho, D. E., Imai, K., King, G. & Stuart, E. A. (2007) Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis*, 15, 199-236.

- Jensen, C. B. (2007) Implementing Europe: A Question of Oversight. *European Union Politics*, 8, 451-477.
- Kaeding, M. (2006) Determinants of Transposition Delay in the European Union. *Journal of Public Policy*, 26, 229-253.
- 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.
- Kaeding, M. (2008a) Lost in Translation or Full Steam Ahead: The Transposition of EU Transport Directives across Member States. *European Union Politics*, 9, 115-143.
- Kaeding, M. (2008b) Necessary conditions for the effective transposition of EU legislation. *Policy And Politics*, 36, 261-281.
- King, G., Keohane, R. & Verba, S. (1994) *Designing Social Inquiry*, Princeton, New Jersey, Princeton University Press.
- König, T. & Luetgert, B. (2009) Troubles with Transposition? Explaining Trends in Member-State Notification and the Delayed Transposition of EU Directives. *British Journal of Political Science*, 39, 163-194
- Lampinen, R. & Uusikyla, P. (1998) Implementation Deficit - Why Member States do not Comply with EU Directives? *Scandinavian Political Studies*, 21, 231-251.
- Linós, K. (2007) How Can International Organizations Shape National Welfare States? Evidence From Compliance With European Union Directives *Comparative Political Studies*, 40, 547-570.
- Luetgert, B. & Dannwolf, T. (2009) Mixing Methods: A Nested Analysis of EU Member State Transposition Patterns. *European Union Politics*, 10, 307-334.
- Mastenbroek, E. (2003) Surviving the Deadline. The Transposition of EU Directives in the Netherlands. *European Union Politics*, 4, 371-395.
- Mastenbroek, E. (2005) EU compliance: Still a 'black hole'? *Journal of European Public Policy*, 12, 1103-1120.
- Mastenbroek, E. (2007) *The politics of compliance. Explaining transposition of EC directives in the Netherlands*, Wageningen, Ponsen & Looijen.
- Mbaye, H. (2001) Why National States Comply with Supranational Law? Explaining Implementation. Infringements in the European Union, 1972–1993. *European Union Politics*, 2, 259-281.
- McKim, V. & Turner, S. (Eds.) (1997) *Causality in Crisis? Statistical Methods and the Search for Causal Knowledge in the Social Sciences*, Notre Dame, University of Notre Dame Press.
- Perkins, R. & Neumayer, E. (2007) Do Membership Benefits Buy Regulatory Compliance?: An Empirical Analysis of EU Directives 1978—99 *European Union Politics*, 8, 180-206.
- Sedelmeier, U. (2009) Post-accession compliance with EU gender equality legislation in post-communist new member states. *European Integration Online Papers (EIoP)*, 13.
- Siegel, S. (2006) Law and Order in the European Union: Explaining Variations in Compliance with European Union Law. *WZB Discussion Paper*, IV, 1-30.
- Steunenberg, B. (2007) A policy solution to the European Union's transposition puzzle: Interaction of interests in different domestic arenas *West European Politics*, 30, 23 - 49
- Steunenberg, B. & Kaeding, M. (2009) 'As time goes by': Explaining the transposition of maritime directives. *European Journal of Political Research*, 48, 432-454.
- Steunenberg, B. & Toshkov, D. (2009) Comparing transposition in the 27 member states of the EU: the impact of discretion and legal fit. *Journal of European Public Policy*, 16, 951 - 970.

- Thomson, R. (2007) Time to Comply: National Responses to Six EU Labour Market Directives Revisited. *West European Politics*, 30, 987 - 1008.
- Thomson, R. (2009) Same effects in different worlds: the transposition of EU directives. *Journal of European Public Policy*, 16, 1-18.
- Thomson, R., Torenviold, R. & Arregui, J. (2007) The Paradox of Compliance: Infringements and Delays in Transposing European Union Directives. *British Journal of Political Science*, 37, 685-709.
- Toshkov, D. (2007a) In search of the worlds of compliance: culture and transposition performance in the European Union. *Journal of European Public Policy*, 14, 933-954.
- Toshkov, D. (2007b) Transposition of EU Social Policy in the New Member-States. *Journal of European Social Policy*, 17, 335-348.
- Toshkov, D. (2008) Embracing European Law: Transposition of EU directives in Central and Eastern Europe. *European Union Politics*, 9, 379-342.
- Toshkov, D. (2009) *Between Politics and Administration. Compliance with EU Law in Central and Eastern Europe*, PhD Dissertation. Leiden University.
- Treib, O. (2006) Implementing and complying with EU governance outputs. *Living Reviews in European Governance*, 1.
- Zhelyazkova, A. & Torenviold, R. (2009) The Time-Dependent Effect of Conflict in the Council on Delays in the Transposition of EU Directives. *European Union Politics*, 10, 35-62.