





Forecasting candidate states' compliance with EU accession rules, 2017–2050

Tobias Böhmelt & Tina Freyburg

To cite this article: Tobias Böhmelt & Tina Freyburg (2017): Forecasting candidate states' compliance with EU accession rules, 2017–2050, Journal of European Public Policy, DOI: [10.1080/13501763.2017.1348385](https://doi.org/10.1080/13501763.2017.1348385)

To link to this article: <http://dx.doi.org/10.1080/13501763.2017.1348385>

 View supplementary material 

 Published online: 12 Jul 2017.

 Submit your article to this journal 

 View related articles 

 View Crossmark data 



Forecasting candidate states' compliance with EU accession rules, 2017–2050

Tobias Böhmelt^{a,b} and Tina Freyburg^c

^aDepartment of Government, University of Essex, Colchester, UK; ^bSwiss Federal Institute of Technology (ETH), Zurich, Switzerland; ^cSchool of Economics and Political Science, University of St Gallen, St Gallen, Switzerland

ABSTRACT


The European Union (EU) is said to be tired of enlargement – but how likely is it that a candidate would be ready to join within 10, 15 or more years? This research forecasts how prospective members are likely able to perform in implementing EU law until 2050. Using compliance data of all states from the 2004, 2007 and 2013 accession rounds, as well as of five current/potential candidates, we construct an empirical model on candidates' ability to comply with the *acquis communautaire*. We employ in-sample and out-of-sample techniques to ensure high model prediction accuracy and, ultimately, forecast the five candidates' potential compliance levels in 2017–2050. Our research shows that *only one* candidate might sufficiently be able to comply with the accession criteria until 2023, while many are unlikely to be ready before the mid-2030s. Focusing on prediction and forecasting, our contribution is given by the research's policy relevance and its methodological innovation.

KEYWORDS Accession conditionality; *acquis communautaire*; European Union; forecasting; prediction

'Enlarging the EU has taken a back seat as members grapple with problems closer to home.'¹ What seems a contemporary statement about European Union (EU) enlargement in times of crisis is the *Financial Times* editor, Lionel Barber, writing in the mid-1990s. Although the EU was generally committed to opening its gates to the East back then, it was careful not to encourage the Central and East European countries to push for membership (Mattli and Plümpner 2002). Indeed, Barber goes on by quoting a senior Commission official saying the '[EU's current] level of seriousness about enlargement is not minimal, it simply does not exist.'

This episode illustrates that enlargement historically stems from the pressure of countries aspiring to join, not from an expansionist ambition on the EU's side. It emphasizes that enlargement and its potentially destabilizing effects have been subject to heated debates since the EU's formation

CONTACT Tina Freyburg  tina.freyburg@unisg.ch

 Supplemental data for this article can be accessed at <https://doi.org/10.1080/13501763.2017.1348385>.

© 2017 Informa UK Limited, trading as Taylor & Francis Group

(O'Brennan 2014: 223f). And yet, parallel to its institutional deepening, the EU grew from originally 6 members to 28 by 2017. This raises key analytical and empirical questions: what would be an adequate baseline model for predicting candidate countries' ability to comply with EU law? And, against previous and ongoing enlargement experiences, how likely is it that a candidate would be ready or able to join over the next decade(s)?

EU enlargement policy is path-dependent. While each member state has, in theory, the possibility to veto at multiple stages, the self-reinforcing nature of a series of sequenced decisions makes it rather difficult to reverse (Giandomenico 2009; Vachudova 2007). Its institutional architecture, owing to the EU's hybrid nature, adds to this rigidity as enlargement requires the approval of central EU institutions, notably the European Parliament and the Commission, which are committed to 'the firm prospect of EU membership' (European Commission 2016). The standardized procedure obliges them to consider all applications from 'European' countries according to the same standards. This rigid architectural design of enlargement policy allows us to presume a certain continuity against which we can assess a candidate's ability to comply. That is, if EU enlargement policy is generally consistently linked to compliance in the target countries (Schimmelfennig 2008) and it is, then and primarily, properties of candidates that explain variance in compliance ability, models on candidates' past compliance can, in principle, accurately predict and forecast their future ability to comply.

Drawing on previous statistical accounts of candidate countries' compliance with EU accession requirements (Böhmelt and Freyburg 2013, 2015; Hille and Knill 2006; Schimmelfennig and Scholtz 2008; Toshkov 2008), we consider a set of core exogenous predictors to construct a baseline model for the forecast. We examine the predictive power of this model via in-sample and out-of-sample techniques. Finally, after having demonstrated our model's prediction accuracy and applicability, we forecast five EU candidates' (Albania, Bosnia-Herzegovina, the Former Yugoslav Republic of Macedonia, Serbia and Turkey) compliance levels with the EU *acquis communautaire* until 2050.

Previous studies developed and tested influential theoretical frameworks that explain the mechanisms of candidates' compliance with the EU *acquis*. However, the empirical evidence is often conducted *ex post* on observed data. Despite important insights, there are limitations associated with these kinds of studies: drawing inferences based on statistical significance testing alone might be misleading about the (predictive) power of an empirical model (Schneider *et al.* 2010; Ward *et al.* 2010: 364). While statistically significant results may improve our understanding of the relationship between variables in each sample under study, they cannot provide information on the exact same relationship in another, i.e., new sample of data – like the future. Prediction and forecasting methods can help in this regard.

Following Hegre *et al.* (2017: 114), we define

forecasts as predictions about unrealized outcomes given model estimates from realized data. ... 'Prediction' is a more general concept, and refers to the assignment of a probability distribution to an outcome based on such model estimates, but may be applied to realized as well as unrealized outcomes. More colloquially, forecasts are predictions about tomorrow given information we have about what has happened up to today. This means two inputs are required to make forecasts: realized data and estimators; and one output is produced: predictions.

Therefore, we consider prediction and forecasting techniques as valuable, in both scientific and practical terms. Overall, we seek to contribute to both the academic and the policy-oriented literature. First, our work provides an assessment when current and potential candidates might be ready for EU membership in light of their ability to adjust to the *acquis*, if at all, through systematic research based on information on candidates' previous levels of compliance. Since earlier work has paid little attention to predicting and forecasting states' ability to implement EU law, policy-makers lack guidance for assessing the success of EU enlargement politics and, hence, making an informed statement on potential readiness of candidates for accession.

Second, we develop a model that enables forecasts of candidates' compliance levels, i.e., their ability to comply with the *acquis* that future research can easily rely on, verify and extend. Specifically, we predict future compliance levels after having determined that our set of exogenous factors (including fixed effects) predicts accurately observed compliance patterns with genuine in-sample and out-of-sample techniques. However, neither is the predictive model in t based on data in t , nor do we use earlier compliance patterns to forecast future compliance. To this end, we provide a comprehensive discussion of our conceptualization of compliance, the specifications of the explanatory variables identified in earlier work, and, most crucially, previous findings. As we provide clear guidelines for prediction exercises in general, we seek to contribute to the 'forecasting literature' in EU politics (Bechtel and Leuffen 2010; Thomson and Hosli 2006) and to enrich the debate on the validity of policies based on empirical models (Schneider *et al.* 2010; Ward *et al.* 2010).

Using the 2004 accession round as a benchmark, the results show that *only one country* of the current and potential EU candidates is likely to be able to sufficiently comply with the accession criteria until 2023, Macedonia, while most of them may not be ready for accession before the mid-2030s. Our forecast highlights that Albania or Bosnia-Herzegovina may even face difficulties in satisfactorily incorporating EU legislation before 2050. Note, however, that this forecast captures the ability of states to comply with the EU *acquis*. We do not forecast actual accession dates or states' willingness – or unwillingness – to comply with EU law. Still, our results underline that populist foreign policy

positions such as the British Vote Leave campaign's claim that most of the candidate countries would join by 2020 seem mistaken.²

Path dependency in EU enlargement and candidates' compliance

Enlargement policy is portrayed as a policy that, once established, is increasingly difficult to reverse, despite member states having potentially deviating preferences. In fact, the foundation of EU enlargement policy remained mostly unchanged since the 1990s. Modifications so far typically concerned minutiae, motivated by candidate specificities rather than EU-internal politics (Schimmelfennig 2008). It appears that as the process unfolds, and a country is 'administratively put under the responsibility of DG Enlargement' (Giandomenico 2009: 111), an ultimate rejection of accession becomes increasingly unlikely (Böhmelt and Freyburg 2013: 267). Hence, amidst economic and political crises, enlargement policy appears to be 'running on autopilot' (Vachudova 2014: 123). Considering the remarkable continuity of enlargement policy and its reactive nature, revolutionary change seems unlikely.

Naturally, the idea of path dependency does not preclude the possibility of institutional or policy changes following an exogenous shock, provided they are sequenced correctly (Thelen and Steinmo 1992). However, anecdotal evidence suggests that significant contemporary events might have no such effect. For example, in July 2013, Croatia became a member of the Union; in January 2014, Serbia's accession negotiations formally began; in February 2016, Bosnia-Herzegovina submitted its application to join the EU; and at the end of June 2016, negotiations on a new chapter were opened with Turkey. Overall, it seems reasonable to assume that future enlargement policy will continue to develop at roughly the same historical pace. What would be an adequate baseline model for explaining and, more importantly, *predicting* candidate countries' ability to comply?

The EU's pre-accession policy is a demanding exercise for any candidate. As determined by the Copenhagen European Council in 1993, states aspiring to join the EU must bring myriad domestic laws, regulatory frameworks and administrative practices (see Endnote 5) in line with the *acquis* prior to an ultimately uncertain accession decision. However, when opening accession negotiations, the EU creates an expectation that 'the applicant country will, at some yet unknown moment join' (Steunenberg and Dimitrova 2007: 3), once it complies with the *acquis*. But under what conditions are candidates likely to be able to formally adopt and then apply EU rules and regulations?

Existing quantitative research³ on candidates' compliance with EU accession requirements vary not only regarding the conceptualization and operationalization of the dependent variable, but also regarding the explanatory variables included and, most importantly, the findings. This work can be

grouped into studies centring on the EU's effect on the quality of democracy in candidate states (Schimmelfennig and Scholtz 2008) and those seeking to explain variation in compliance levels with the *acquis* prior to accession (Böhmelt and Freyburg 2013, 2015; Hille and Knill 2006; Toshkov 2008). We focus on the latter strand, especially the work by Hille and Knill (2006), as well as Böhmelt and Freyburg (2013, 2015), since they share similar research interest and the same compliance measurement (discussed below). Adopting a rationalist perspective, complemented with managerialist insights, these studies concentrate on a country's relevant domestic actors' ability, and also partly their willingness, to adjust national legislation to EU rules.

Despite the substantive similarities, existing studies vary regarding research designs and findings. The empirical analyses differ in country years, with Hille and Knill (2006) studying 13 candidates in 1999–2003 and Böhmelt and Freyburg (2013, 2015) covering an extended sample of 16 states in 1998–2009. Furthermore, Böhmelt and Freyburg (2013) employ generalized additive models, while a linear model is used in both Hille and Knill (2006) and Böhmelt and Freyburg (2015). Most covariates are measured with the same data; yet, while Böhmelt and Freyburg (2013) use the World Bank Development Indicators to capture a bureaucracy's financial capabilities, Hille and Knill (2006) rely on data provided by the CIA World Factbook for government expenditures and use the Commission's regular reports for gross domestic product per capita. It seems that the extent to which the relationship between candidates' levels of compliance and the theoretically derived determinants of effective conditionality is statistically significant in regression models considerably varies, too (see Online Appendix 1).⁴

All three studies confirm that conditionality is the main force driving candidates' compliance with the EU *acquis*. However, Hille and Knill (2006: 549) demonstrate that 'the functioning and the quality of the domestic bureaucracy constitute crucial preconditions.' Böhmelt and Freyburg (2015) add that candidate states may free-ride on the compliance efforts of others, and that, independently from country and EU-level conditions, the leverage of conditionality varies over the pre-accession process. The question remains whether existing explanatory models are also suited to *predict* a candidate's compliance ability.

Again, results in the form of regression coefficients may tell little about how candidates' compliance with EU law will develop in the future. Policy prescriptions cannot be based on statistical summaries of probabilistic models. Hypothesis testing that ignores out-of-sample heuristics faces the inherent risk of fitting to a specific sample's idiosyncrasies rather than identifying stable structural relationships between an outcome of interest and its determinants. In fact, if a model *explains* the relationship between, in our case, states' ability to comply with EU law and some explanatory factors well in-sample, we merely assume that it also performs well when presented with

new data and trying to predict *out-of-sample*. But if the model only gives a description of this relationship in the original data set without capturing underlying causal relations, the chances to make correct predictions with new data are likely to be undermined.

Our analysis seeks to address this shortcoming. We first identify the necessary variables for our model and then predict candidates' ability to comply with EU law in-sample and out-of-sample. Moving from empirical analyses based on statistical significance to prediction offers a more solid scientific basis for assessing future levels of compliance with EU accession law, which is highly relevant both from a policy and scholastic perspective (Hegre *et al.* 2017; Schneider *et al.* 2010). It also allows us to analyse whether our selected model based on in-sample prediction can accurately predict candidates' ability to comply out-of-sample, as theoretically suggested by the continuity argument.

Research design

Empirical strategy and dependent variable

Our predictions and forecasts are based on an OLS regression model analysing data on candidates' compliance with EU law using core exogenous predictors, which are meant to primarily proxy the costs and capabilities arguments emphasized in previous empirical studies. Our sample consists of 18 (previous, current and potential) candidates for which reliable data are available (see Online Appendix 2). We use the country-year as the unit of analysis, with a state dropping out of the sample once the Commission no longer provides progress reports; this happens either at the time when accession treaties are signed or with the latest reports covered by this study (i.e., European Commission 2016).

To arrive at our forecast of states' compliance ability in 2017–2050, we complete the following steps (see Online Appendix 5). First, we identify a set of potential predictors that may help us in explaining candidates' ability to comply and, thus, their compliance levels. A crucial requirement for these predictors is not only that they are exogenous to our dependent variable, but also have available high-quality projections until 2050. Not many variables meet these criteria. We follow Hegre *et al.* (2013) and focus on a set of socio-demographic variables for which projections until 2050 exist and that can be linked to the quantitative literature on effective EU enlargement policy and, especially, previous claims about costs and capabilities. In addition, we add fixed effects for enlargement rounds and states, which capture temporal shocks, EU policies specific for each enlargement rounds or unobserved unit-level influences affecting compliance.

Second, for determining the prediction accuracy of this model, we estimate it on a time-series cross-sectional sample in 1998–2008 ($N = 120$ country

years), which we then assess with in-sample techniques. Third, we examine the out-of-sample prediction power by employing a four-fold cross-validation exercise and comparing our predictions for 2009 to 2016 (based on the estimates for the 1998–2008 period) with observed values (for which we have data). After having assessed and confirmed the predictive power of our model the final forecast for the 2017–2050 period is based on a model that uses the entire set of already observed data in 1998–2016. We opted for 2050 as the final year in our forecast, as per Hegre *et al.* (2013), given we have only 19 years of observed data points. We present results for the aggregated sample as well as individual countries.

To quantify and measure candidates' compliance with EU law, our dependent variable, we employ the updated compliance-level data from Böhmelt and Freyburg (2013, 2015), who use the EU Commission's annual reports⁵ on each candidate's progress in aligning policies toward EU requirements. In these reports, the Commission explains and assesses in detail what each candidate has achieved over the last year, and identifies areas where more effort is needed to have the 'ability to assume the obligations of membership'.

Reporting on a candidate stops the year the accession treaty was signed for the 2004 accession-round states (2003). For Bulgaria and Romania, progress reports are given until 2005 only, but we cover the last year (2006) before the 2007 accession with a joint report.⁶ Similarly, the Commission published a 'comprehensive monitoring report on Croatia's state of preparedness for EU membership' in 2012 (i.e., one year after the accession treaty has been signed), which follows the same structure as the annual progress reports; we use this file to code the last year before Croatia's accession and include this country-year in our data set as well. The Commission reports have the advantage that their data quality is high and that they evaluate both formal and practical compliance with EU law of each candidate state on an annual basis in a standardized and comparable manner (Hille and Knill 2006: 541f).

The final dependent variable, a country's logged degree of compliance with EU law in each policy area, is coded along the ordinal four-value assessment provided by the Commission (Böhmelt and Freyburg 2013, 2015): the value of 0 is assigned when a country does not comply with the *acquis* in a specific issue area; 1 if a country partly complies with EU laws and regulations in a specific issue area, although substantially more efforts are necessary; 2 if a country almost fully complies with the *acquis* in a specific issue area, although more efforts are necessary; and 3 when a country fully complies with EU laws and regulations in a specific issue area. Each sector thus receives a value between 0 and 3, while higher values signify higher compliance with the *acquis*.

Böhmelt and Freyburg (2013, 2015) then estimate the average degree of (logged) compliance for a country in each year by calculating the mean value across all policy areas plus the general evaluation and taking the

natural logarithm. Focusing on the ‘more general rather than issue or policy-specific’ (Hille and Knill 2006: 535) performance of countries, this strategy ensures that we receive a standardized, hence comparable, measure for all countries at different enlargement stages and rounds. Finally, we include the 2016 Commission reports, which comprise compliance patterns that are likely to be affected by populist backlashes of nationalist and populist Euroscepticism, including the arguably most extreme change in candidate countries: the authoritarian turn in Turkey that militates against meeting the EU’s political accession conditions.

Predictors

We specify a model with core explanatory variables that fulfil three pivotal forecasting criteria. First, the chosen variables are exogenous to our ‘indirect measure’ (Toshkov 2008) of compliance performance with EU accession rules based on Commission reports (or they are time-invariant). Second, they arguably proxy the costs and capabilities arguments emphasized by existing statistical studies. Finally, good projections until 2050 are available for the time-variant items. The selected variables are a time trend, fixed effects for enlargement rounds, country-fixed effects, a state’s demographic composition, infant mortality and education (see Online Appendix 3).

Commonly used operationalizations and variable specifications in earlier studies may suffer from possible endogeneity with our outcome variable, candidates’ compliances as reported by the Commission, for predominantly two reasons. First, the Commission stresses the need for administrative and judicial capacity to ensure correct implementation and application of the many rules next to the actual adoption of the *acquis* (Christoffersen 2007: 47). Hence, dependent and core explanatory variables in previous work might conceptually overlap. Second, the expert scores used to measure some determinants of candidates’ compliance might inform the Commission’s assessment of compliance with its accession rules or have been informed by its progress reports (Toshkov 2008: 382). For instance, state capacity is frequently measured by the expert ratings collected for the World Bank; political costs or incentives are often operationalized with data on a country’s level of democracy from the Polity IV project. The Commission uses many sources including contributions from the respective candidate government, the member states, the European Parliament reports, as well as information from various international and non-governmental organizations (Christoffersen 2007: 31). At the same time, the Commission makes ‘efforts to ensure that international organizations such as IMF and World Bank pay attention to the reports’ (Kelley 2006: 34).

Moreover, for commonly used operationalizations, notably the anticipated adoption costs based on a country’s level of political and economic liberalization or its capabilities in terms of bureaucratic strength or government

expenditures, reliable projections are not available for the period 2017–2050. Therefore, we use projections for predictors – demographics, infant mortality and education – from the United Nations (UN) World Population Prospects and the International Institute for Applied Systems Analysis (IIASA) that we believe can serve as proxies for commonly suggested covariates. The World Population Prospects provides ‘estimates of demographic indicators for all states in the international system’ (Hegre *et al.* 2013: 254) and projections for these key variables until 2050. These projections are based on national population censuses and revised in biennial consultation with experts from national projection-making agencies. We take the mean scenarios of the UN projections and complement them with those provided by the non-governmental research organization IIASA based on expert and argument-based forecasting, in particular its 2001 revisions of the World Population Program, as released in its final form in 2004.⁷ Despite inherent demographic uncertainty, the estimates of demographic indicators provided by both UN and IIASA are seen as the most authoritative (Hegre *et al.* 2013; O’Neill *et al.* 2001: 206).

Our first cluster of predictors seeks to proxy the domestic costs associated with compliance by a target government. A government’s preferences are likely to be influenced by the extent to which the public supports EU membership and, hence, the likelihood of the public to punish potentially costly reforms in coming elections.⁸ Various studies on the relationship between societal characteristics, mass attitudes and EU integration/membership respectively emphasize socioeconomic determinants of EU support, with some groups gaining and others losing from membership. From this utilitarian perspective, an individual’s expected net gain from EU membership significantly depends on her individual characteristics, notably *age* or *education*. While studies of attitudes in the then-current EU members consistently demonstrate that ‘winners’, commonly the young and better educated, are more likely to have favourable EU attitudes (Gabel 1998), the evidence for such individual-level drivers of EU support in candidate countries remains mixed. In fact, studies relying on the Central and Eastern Eurobarometer survey data find only weak and cross-nationally inconsistent effects of demographic characteristics on support for EU membership (Tucker *et al.* 2002: 569; Tverdova and Anderson 2004). Work using alternative data sources, namely the actual results of the referendums at the regional level, however, find that higher level of education increases the likelihood of voting in favour of EU membership (Doyle and Fidrmuc 2006). We include demographic variables in our analysis, measuring age and education. The education data are taken from IIASA, which uses definitions and categories consistent over countries and time to facilitate cross-national and time-series comparisons (Lutz *et al.* 2007). Precisely, we employ a measure of male secondary education, defined as ‘the proportion of men aged 20–24 years with secondary or

higher education of all men aged 20–24' (Hegre *et al.* 2013: 255).⁹ The age data (*Youth*), in turn, come from the UN World Population Prospects series providing age-specific population numbers 'measured as the percentage of the population aged 15–24 years of all adults aged 15 years and above.'

As to the capabilities argument, that is a state's administrative capacity to adopt and implement EU rules (or its ability to implement adjustment efforts at the domestic level), factors such as economic development and regime type are explanatory variables commonly incorporated in probabilistic models of effective EU conditionality (Böhmelt and Freyburg 2013, 2015). However, as discussed above, these factors do not meet this study's requirements. Instead, quantitative researchers commonly measure state capacity by looking at the outcomes of public goods and service delivery, such as the percentage of children enrolled in primary schools, infant mortality rates or literacy rates. These measures are attractive owing to their broad coverage and cross-national comparability, although not without criticism (Hanson 2015), because they can be attributed to factors other than state capacity, including levels of economic development and political regime type. Since we do not test the costs and capabilities arguments against each other, but use them to identify a prediction model for candidates' future ability to comply with EU law, we do not perceive this a problem for our study. Considering this discussion, among the traditional measures of state capacity in terms of administrative performance (Bäck and Hadenius 2008), infant mortality rates present a particularly useful composite indicator of the provision of public services. Taking the data from the UN (2007), infant mortality is defined as the probability of dying between birth and exact age one year, expressed as the number of infant deaths per 1000 live births.

We also include a time trend correcting for temporal dependencies (Böhmelt and Freyburg 2013, 2015).¹⁰ The EU added more policy areas over the years, which may make compliance more costly, hence less likely. We therefore log-transform the time-related variable *Year*. Our model also incorporates fixed effects for a specific enlargement round (2004, 2007 and 2013) and future enlargement to account for the spatial dependencies and the possibility that the specific requirements became more difficult with each round. These dummies shall further capture any time-invariant group-specific characteristics and unobserved features of each accession round's requirements, as determined by EU policy.

Finally, we include country-fixed effects that are based on the same rationale at the state-level, i.e., capturing unobserved time-invariant unit-level effects that may influence compliance with EU law. In light of these data and methods, particularly the inclusion of country-fixed effects, we utilize a reduced-form approach of estimation and projection, which assumes that owing to its rigid institutional architecture, future EU enlargement policy will continue to develop at roughly the historical pace.

In the following, we demonstrate that the predictors we include can accurately and precisely predict actually observed compliance values, i.e., that our model has little prediction error. Note that the direction of influence or the statistical significance of the covariates does not matter for this purpose (Ward *et al.* 2010): it is the precision and accuracy of our final model in making predictions that counts.

Empirical results

In-sample prediction

To assess the accuracy of the ‘conditional statements about a phenomenon for which the researcher actually has data, i.e., the outcome variable has been observed’ (Bechtel and Leuffen 2010: 311), we first estimate the baseline model in 1998–2008 with ordinary least squares (OLS) (Model 1 in Online Appendix 4), then calculate the predicted values of this model for that time period, and finally compare the predicted yearly median levels of candidates’ compliance using the estimated parameters from the baseline model with the truly observed median compliance between 1998 and 2008. The results are depicted in the left panel of Figure 1.

While the dashed line captures our predicted values as derived from the parameters of our model, the solid line pertains to the observed values of compliance with the EU *acquis*. The left panel in Figure 1 shows that the model slightly over-predicts compliance values until about 2000. At the same time, predicted and observed values are *almost identical as of 2003*, except for the curve’s tail, where our model marginally under-predicts compliance. The decline in compliance from 2003 to 2004 is driven by those states

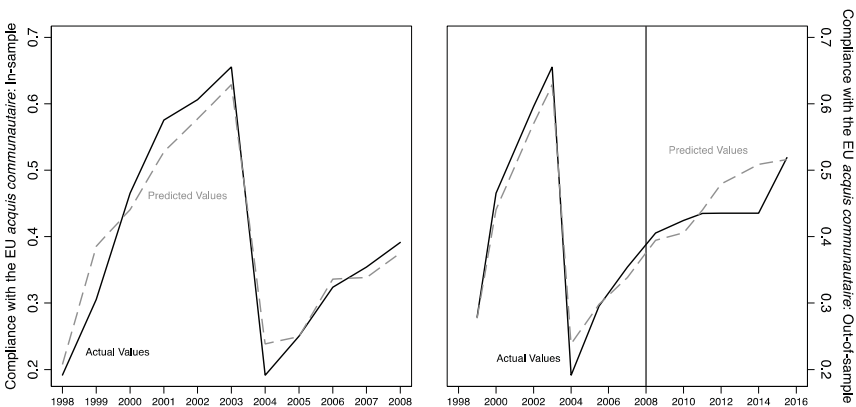


Figure 1. Median levels of compliance with EU accession rules – in-sample (left) and out-of-sample (right) prediction.

Note: Predicted values are indicated by the dashed line and actual values by the solid line.

joining the EU in 2004, which drop out of the sample (as the Commission ceases to provide progress reports with accession); the 'remaining' countries in the sample have been, as they were not ready for joining the EU back then, characterized by significantly lower compliance scores. Overall, this graph demonstrates that the predicted values fit the time points of the observed data extremely well.

To assess the accuracy of this prediction more thoroughly, we use two goodness-of-fit measures: the mean squared prediction error (MSPE) and Theil's U (Theil 1966), which (unlike the MSPE) does not depend on the scale of the data (Bechtel and Leuffen 2010). Theil's U is the square root of the ratio between the sum of squared prediction errors of the baseline model (Model 1 in Online Appendix 4) and the sum of squared prediction errors of a naïve model, i.e., a 'no-change prediction' where the level of candidate compliance with EU law in $t-1$ fully corresponds to the level of compliance in t . If Theil's U is larger than 1, the model performs worse than the naïve model; values for Theil's U smaller than 1 indicate that the 'theoretically informed model' performs better than the naïve specification. The closer the MPSE is to 0, the more accurate is the model in making predictions. For our model, the MPSE is 0.0065, while Theil's U stands at 0.723.

Overall, the specification used in the model performs well in predicting candidate states' compliance with EU law *in-sample*. It remains to be seen how accurately this model predicts candidate states' compliance when moving to the 'harder' test of an out-of-sample prediction confronting the model with 'new' data. What is the model's predictive power when trying to correctly predict compliance that is not 'within the very same set of data that was used to generate the models in the first place' (Ward *et al.* 2010: 8)?

Out-of-sample prediction

For the out-of-sample prediction, we first use a four-fold cross-validation quasi-experimental setup that was repeated 10 times for the baseline model in 1998–2008 (Ward *et al.* 2010: 370). Cross-validation randomly divides our sample we employed for the baseline model into four segments. We use three segments to estimate the parameters, while the fourth, 'test-set' segment is retained for assessing the predictive power of the baseline model on the pooled subsets. Therefore, there are three segments of the data to build the model and create predictions. The remaining (randomly chosen) segment is not considered for estimating the model but treated as if 'unknown'; its mere purpose is for comparing the predicted with the observed values. Again, we calculated the MPSE and Theil's U for the predictive power, for which we then present the average values over the 10 repetitions. The average MSPE for the cross-validation exercise stands at 0.012, while Theil's U has an average value (across the 10 repetitions of the exercise) of 0.966 now.

We graphically depict predicted and actual values of candidates' compliance with EU law in the right panel in [Figure 1](#). The difference between the two panels in [Figure 1](#) is that the right panel extends the period of study to 2016, as we compare the observed values of candidates' compliance with EU law in 1998–2008 (our 'observed' time period) with those 'unobserved' values in 2009–2016, i.e., the time period that we have not used for building the model and that we treat as 'unobserved', although we know the true values.

Two conclusions can be derived. First, uncertainty remains and, unsurprisingly, the predictions for the 'unobserved' data partition are less accurate than in the case of the in-sample prediction. Two indicators demonstrate this: (1) our goodness-of-fit measures, which both increase and, therefore, show that prediction power decreases;¹¹ and (2) the comparison between observed (1998–2008) and predicted values (2009–2016) in [Figure 1](#). Second, prediction accuracy is still strong, even when confronting the model with new data. Hence, we move on to the core contribution of this article: the out-of-sample forecast of candidates' compliance with EU law in 2017–2050 for individual countries and aggregated median predictions.

Out-of-sample forecast

The underlying model we use for the forecast is fully based on Model 1 (see [Online Appendix 3](#)) with one exception: we no longer restrict the time period used for building the model's parameters to 1998–2008 but use the entire time period our dependent variable has data for, i.e., 1998–2016. The estimates of Model 2 (see [Online Appendix 3](#)) are virtually identical to those in Model 1. However, the relevant question is how this model predicts the five current and potential candidate states' compliance levels with the EU *acquis* for the future?

To this end, we calculated the predicted values for *EU Accession Compliance* in 1998–2050. These calculations are partly based on data that helped building the model, i.e., we use data that cover 1998–2016. However, all data points after 2016, i.e., 2017–2050, are not part of our model as we do not have observed values for candidates' compliance here. This constitutes the true forecast. We plot these predicted values next to the observed values in the upper-left panel of [Figure 2](#). The solid line signifies the observed values (in 1998–2016), while the two vertical solid lines mark the points in time at which we set thresholds for the in-sample and out-of-sample predictions before (i.e., 2008 and 2016). We capture the uncertainty inherent in our forecast by including upper and lower bound confidence intervals for the predictions (upper and lower dashed lines in the upper-left panel of [Figure 2](#)).

We also calculated the predicted values for each of the five current or potential candidate states individually, see [Figure 2](#).¹² The horizontal solid

line in these country panels pertains to a reasonable benchmark we set for ‘sufficiently high’ compliance. Schimmelfennig and Sedelmeier (2004: 666) acknowledge that the EU might undermine the credibility of its conditionality if it admits candidates at different levels of preparedness. Testimonies of accession negotiations point out that enlargement decisions have been considered for a group of countries as a whole, e.g., the Baltic States or the Visegrad countries rather than for each state individually. Particular groups may join when a sufficiently large number of candidates shows good compliance (Christoffersen 2007: 32f). We thus assume that all states entering the EU in 2004 (1) had achieved more or less the same level of ‘preparedness’ so that they could jointly assume membership, and (2) that this joint level of compliance was also sufficiently high. We use the average level of compliance with EU law of all states that joined the EU in 2004 as a benchmark value for future accessions – this benchmark lies at 0.733, according to our data.¹³ Note that this forecasts states’ ability to comply not their willingness; we do not predict actual accession dates (but years in which candidate states might, in principle, be able to comply with EU law).

Several important conclusions can be derived from Figure 2. First, the high prediction power we identified in our model in the previous sections remains to be given. This is illustrated by comparing the predicted and ‘real’ values in 1998–2016 (upper-left panel). Second, overall ability to comply with EU law is

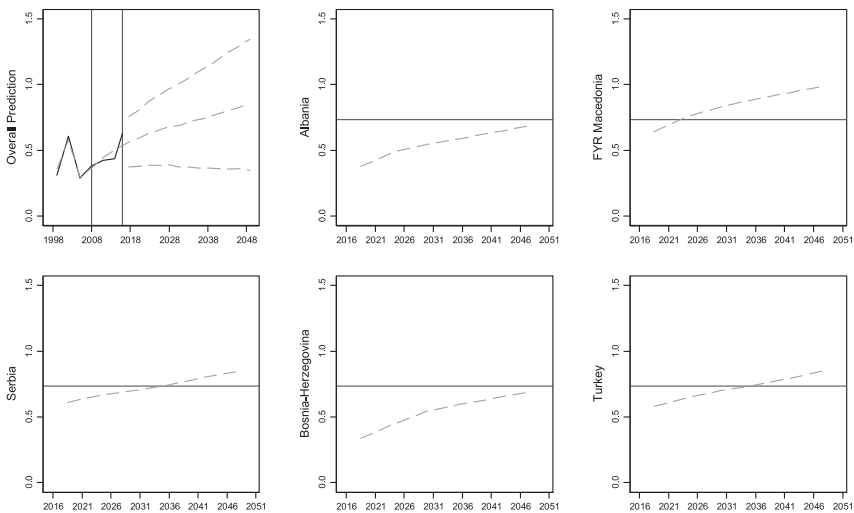


Figure 2. Median levels of compliance with EU *Acquis*.

Notes: Predicted values are indicated by the dashed line and actual values by the solid line. Solid horizontal line marks benchmark compliance value (2004 accession). Vertical solid lines in upper-left panel pertain to the points in time at which we set thresholds for the in-sample and out-of-sample predictions (i.e., 2008 and 2016). Upper and lower bounds of 90 per cent confidence interval are included in upper-left panel in 2017–2050 as well.

supposed to increase over time, regardless of which scenario we look at. Third, and most interestingly, compliance ability varies substantially by country.

Despite this variation, though, *only one* of the current/potential EU candidates seems to be able to sufficiently comply with the *acquis* until 2023, while most are unlikely to be ready for accession before the mid-2030s when taking the 2004 average compliance level as a benchmark. By 2023, only Macedonia is likely to meet the 2004-accession benchmark. Serbia passes the threshold in 2035, Turkey – even after accounting for the recent events pertaining to a democratic backlash, as reported in the 2016 Commission report – is likely to have the ability to sufficiently comply in 2036, while Bosnia-Herzegovina (2050: 0.715) and Albania (2050: 0.722) might find it difficult to meet the set standard even by 2050.

Conclusion

EU candidates are required to adjust domestic legislation prior to accession for bringing their laws, regulations, and administrative practices in line with the *acquis communautaire*. Knowing in advance which of the (potential) current candidates are less able to abide by EU regulations over the course of accession is not only of academic interest, but also essential to the EU's monitoring and enforcement schemes as well as an informed public debate about future EU enlargement. Yet, thus far, we knew relatively little about the actual accession prospects of current candidates, in particular how their compliance ability may develop over the years to come. Previous empirical testing is primarily of an *ex-post* nature, thus merely accounts for compliance patterns in the past. We take research on candidates' compliance with the EU accession rules one step further by moving from *ex-post* analysis to predictions and forecasts about likely future compliance.

We specified a model to predict the ability to comply with the EU *acquis* of potential/current candidates based on proxies for adjustment costs and administrative capacities, and for which we have observations back to 1998 and projections up to 2050. We used in-sample and out-of-sample techniques to assess our model's predictive power, before providing out-of-sample forecasts of candidates' compliance with EU law in 2017–2050. Our research finds that the empirical model accurately predicts candidates' compliance ability; it emphasizes that *only one country* of the EU's candidates appears able to sufficiently comply with the *acquis* until 2023: Macedonia. Albania or Bosnia-Herzegovina may even face problems in their ability to comply sufficiently before 2050. These figures might lower expectations of further enlargement any time soon.

Note that these figures may still paint 'too optimistic' a picture and the actual future compliance levels of the individual candidate states could well be even weaker than suggested. Our model not only seems to slightly under-predict compliance rates towards the end of the observation period.

Our predictions are also based on some partly restrictive assumptions, namely that the forecasts for our exogenous predictors turn out to be correct, that the past relationship between our predictors and the probability of EU compliance will continue to hold in the future. Also, we cannot account for random events or states' unwillingness to comply in the future, even if they could, such as the current Justice and Development Party (Adalet ve Kalkınma Partisi, AKP) government in Turkey. Nevertheless, we are confident in our predictions.

Assessing the predictive power of empirical models and forecasting future state behaviour have important implications for theory development and can offer significant benefits for policy-makers to foresee candidates' compliance with EU law more accurately. Our study is informative in terms of effectively allocating resources within the EU. The findings suggest that more efforts are necessary if the EU wants its potential future members to better comply with its *acquis*. What is more, if no political decision will be taken favouring early enlargement despite non-compliance, our finding that all candidates except one are unlikely to be able to join within the next 10 years may help to take some heat of the current enlargement debate, which often makes believe that further enlargement is lurking around the next corner.

Notes

1. 'Brussels keeps shut the gates to the East' (Lionel Barber), Financial Times (16/11/1995).
2. <http://www.bbc.com/news/uk-politics-eu-referendum-36355138> (accessed 10 April 2017).
3. Our focus on quantitative research shall not imply that work based on other methods is negligible. Our literature review revealed that existing quantitative research explicitly builds on the key (and statistically testable) insights of the various small/medium-*N* case studies that continue to dominate the field. While the quantitative study of member states' compliance, transposition, and implementation of EU law has flourished (Toshkov 2010), there still are only a few systematic quantitative studies on candidates' compliance in the context of EU enlargement.
4. Online Appendix 1 indicates for all covariates included in the models whether a significant negative (positive), a non-significant negative (positive), or no relationship (~zero) is reported. We distinguish between 'non-significant relationships' and 'no relationship', as statistical significance does not equal substantive significance. In small samples the lack of statistical significance might obscure a substantively important relationship (Toshkov 2010).
5. https://ec.europa.eu/neighbourhood-enlargement/countries/package_en (accessed 10 April 2017).
6. https://ec.europa.eu/neighbourhood-enlargement/sites/near/files/pdf/key_documents/2006/sept/report_bg_ro_2006_en.pdf (accessed 10 April 2017).
7. For more information, see the Guide to Global Population Projections by O'Neill *et al.* (2001). Our additional results based on calculations using the data on other UN scenarios do not qualitatively differ from those reported.

8. An anonymous reviewer suggested that support for EU membership is only one factor regarding adjustment costs and our variables may be more related to perceived legitimacy. Owing to our focus on prediction rather than explanation, the specific relationship a single predictor has with the outcome variable is not really of interest here. What matters is whether this predictor, as part of the full model, contributes to the model's prediction and forecasting power; we demonstrate it does so. Still, though analytically different, the two concepts are related. Strong support might incline a government to accept high adjustment cost, but it does not tell us how high these costs are originally. High support reduces the net costs, but it may not serve as (comparative) indicator of compliance costs.
9. Male and female secondary education should be highly correlated in relatively developed countries in Europe, which is our focus (Breen *et al.* 2010).
10. Auffhammer and Carson (2008: 237) recommend against using year-fixed effects but suggest using a time-trend variable, which is our approach.
11. The corresponding MSPE is 0.010, while Theil's U remains below 1 (0.918).
12. Confidence intervals omitted for the presentation of the point estimates.
13. Using the 2004 accession round as benchmark is based on more data points (more countries) than subsequent rounds, lowering measurement error to some degree.

Acknowledgements

We thank Lena Kiesewetter, Chingun Anderson and Nicolas Cerkez for their research assistance. We are also grateful to Solveig Richter and Frank Schimmelfennig for feedback, while the journal's editors and the anonymous reviewers provided valuable comments that helped us to improve the article. An Online Appendix containing supplemental information on the analyses and additional robustness checks can be downloaded from the author's website (www.tina-freyburg.eu). The corresponding replication material is also freely available from this website.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Tobias Böhmelt is a reader (associate professor) in the Department of Government at the University of Essex, UK, and a research associate at the Centre for Comparative and International Studies (CIS), as well as the Institute for Environmental Decisions (IED) of the Swiss Federal Institute of Technology (ETH) in Zurich, Switzerland.

Tina Freyburg works as professor of comparative politics at the University of St Gallen, Switzerland; she serves as the vice-president of the European Communities Studies Association (ECSA) Switzerland and the co-director of the European Studies working group of the Swiss Political Science Association (with A. Littoz-Monnet).

References

- Auffhammer, M. and Carson, R. (2008) 'Forecasting the path of China's CO₂ emissions using province-level information', *Journal of Environmental Economics and Management* 55(3): 229–47.

- Bäck, H. and Hadenius, A. (2008) 'Democracy and state capacity: exploring a J-shaped relationship', *Governance* 21(1): 1–24.
- Bechtel, M. and Leuffen, D. (2010) 'Forecasting European Union politics: real-time forecasts in political time series analysis', *European Union Politics* 11(2): 309–27.
- Böhmelt, T. and Freyburg, T. (2013) 'The temporal dimension of the credibility of EU conditionality and candidate states' compliance with the *acquis communautaire*, 1998–2009', *European Union Politics* 14(2): 250–72.
- Böhmelt, T. and Freyburg, T. (2015) 'Diffusion of compliance in the 'race towards Brussels?' A spatial approach to EU accession conditionality', *West European Politics* 38(3): 601–26.
- Breen, R., Luijckx, R., Müller W. and Pollak, R. (2010) 'Long-term trends in educational inequality in Europe: class inequalities and gender differences', *European Sociological Review* 26(1): 31–48.
- Christoffersen, P. (2007) 'The preparation of the fifth enlargement', in G. Vassiliou (ed.), *The Accession Story. The EU From Fifteen to Twenty-Five Countries*, Oxford: Oxford University Press, pp. 24–33.
- Doyle, O. and Fidrmuc, J. (2006) 'Who favors enlargement? Determinants of support for EU membership in the candidate countries' referenda', *European Journal of Political Economy* 22(2): 520–43.
- European Commission. (2016) '2016 communication on EU enlargement policy', COM (2016)715 Final, 9 November, Brussels: Commission of the European Communities, available at https://ec.europa.eu/neighbourhood-enlargement/sites/near/files/pdf/key_documents/2016/20161109_strategy_paper_en.pdf (accessed June 2017).
- Gabel, M. (1998) 'Public support for European integration: an empirical test of five theories', *The Journal of Politics* 60(2): 333–64.
- Giandomenico, J. (2009) 'Path dependency in EU enlargement: Macedonia's candidate status from a historical institutionalist perspective', *European Foreign Affairs Review* 14: 89–112.
- Hanson, J. (2015) 'Democracy and state capacity: complements or substitutes?', *Studies in Comparative International Development* 50(3): 304–30.
- Hegre, H., Karlsen, J., Nygård, H., Strand, H., Urdal, H. and Karlsen, J. (2013) 'Predicting armed conflict, 2010–2050', *International Studies Quarterly* 57(2): 250–70.
- Hegre, H., Metternich, N., Nygård, H. and Wucherpfennig J. (2017) 'Introduction: forecasting in peace research', *Journal of Peace Research* 54(2): 113–24.
- Hille, P. and Knill, C. (2006) "'It's the bureaucracy, stupid:" the implementation of the *acquis communautaire* in EU candidate countries, 1999–2003', *European Union Politics* 7(4): 531–52.
- Kelley, J. (2006) 'New wine in old wineskins: promoting political reforms through the New European neighbourhood policy', *JCMS: Journal of Common Market Studies* 44(1): 29–55.
- Lutz, W., Goujon, A., Samir, H.C. and Sanderson, W. (2007) 'Reconstruction of populations by age, sex and level of educational attainment for 120 countries for 1970–2000', *Vienna Yearbook of Population Research* 5(1): 193–235.
- Mattli, W. and Plümpner, T. (2002) 'The demand-side politics of EU enlargement: democracy and the application for EU membership', *Journal of European Public Policy* 9(4): 550–74.
- O'Brennan, J. (2014) "'On the slow train to nowhere?" The European Union, 'enlargement fatigue' and the western Balkans', *European Foreign Affairs Review* 19(2): 221–42.
- O'Neill, B., Balk, D., Brickman, M. and Ezra, M. (2001) 'A guide to global population projections', *Demographic Research* 4(8): 203–88.

- Schimmelfennig, F. (2008) 'EU political accession conditionality after the 2004 enlargement: consistency and effectiveness', *Journal of European Public Policy* 15(6): 918–37.
- Schimmelfennig, F. and Scholtz, H. (2008) 'EU democracy promotion in the European neighbourhood: political conditionality, economic development, and transnational exchange', *European Union Politics* 9(2): 187–215.
- Schimmelfennig, F. and Sedelmeier, U. (2004) 'Governance by conditionality: EU rule transfer to the candidate countries of central and Eastern Europe', *Journal of European Public Policy* 11(4): 661–79.
- Schneider, G., Gleditsch, N.P. and Carey, S. (2010) 'Exploring the past, anticipating the future: a symposium', *International Studies Review* 12(1): 1–7.
- Steunenberg, B. and Dimitrova, A. (2007) 'Compliance in the EU enlargement process: the limits of conditionality', *European Integration Online Papers* 11(5): 1–18.
- Theil, H. (1966) *Applied Economic Forecasts*, Amsterdam: North-Holland Publishing.
- Thelen, K. and Steinmo, S. (1992) *Structuring Politics. Historical Institutionalism in Comparative Analysis*, Cambridge: Cambridge University Press.
- Thomson, R. and Hosli, M. (2006) 'Who has power in the EU? The commission, council and parliament in legislative decision-making', *JCMS: Journal of Common Market Studies* 44(2): 391–417.
- Toshkov, D. (2008) 'Embracing European Law compliance with EU directives in central and Eastern Europe', *European Union Politics* 9(3): 379–402.
- Toshkov, D. (2010) *Taking Stock: A Review of Quantitative Studies of Transposition and Implementation of EU Law*, Leiden: University of Leiden.
- Tucker, J., Pacek, A. and Berinsky, A. (2002) 'Transitional winners and losers: attitudes toward EU membership in post-communist Countries', *American Journal of Political Science* 46(3): 557–71.
- Tverdova, Y. and Anderson, C. (2004) 'Choosing the West? Referendum choices on EU membership in east-central Europe', *Electoral Studies* 23(2): 185–208.
- United Nations. (2007) *World Population Prospects. The 2006 Revision*, New York: UN.
- Vachudova, M. (2007) 'Historical institutionalism and the EU's eastern enlargement', in K. McNamara and S. Meunier (eds), *Making History: European Integration and Institutional Change at Fifty*, Oxford: Oxford University Press, pp. 105–22.
- Vachudova, M. (2014) 'EU leverage and national interests in the Balkans: the puzzles of enlargement ten years on', *JCMS: Journal of Common Market Studies* 52(1): 122–38.
- Ward, M., Greenhill, B. and Bakke, K. (2010) 'The perils of policy by *p*-value: predicting civil conflicts', *Journal of Peace Research* 47(4): 363–75.