



Enforcement tool or strategic instrument? The initiation of ex-post legislative evaluations by the European Commission



European Union Politics

2017, Vol. 18(4) 640–657

© The Author(s) 2017

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1465116517725901

journals.sagepub.com/home/eup



Stijn van Voorst

Tilburg Law School, Tilburg University, Tilburg, the Netherlands

Ellen Mastenbroek

Institute for Management Research, Radboud University, Nijmegen, the Netherlands

Abstract

Whereas the European Commission officially intends to periodically evaluate all major European Union legislation in force, in practice it only evaluates a minority of major regulations and directives. This article tries to explain the variation in the initiation of such ex-post legislative evaluations by the Commission with the help of two theoretical motives: an enforcement motive and a strategic motive. Based on two novel datasets and binary logistic regression analysis, the results show that the type and complexity of the legislation, the presence of an evaluation clause and the evaluation capacity of the responsible Directorates-General enhance the chances of evaluation. These findings indicate that ex-post legislative evaluations are at least partly driven by the Commission's need to enforce legislation.

Keywords

European Commission, legislative evaluation, ex-post evaluation, compliance

Introduction

The European Union (EU) is often described as a 'regulatory state' due to the important role of legislation in the European policy process (Majone, 1999: 1). A marked feature of the European legislative process is the centrality of one

Corresponding author:

Stijn van Voorst, Tilburg Law School, Tilburg University, Tilburg 5038TD, the Netherlands.

Email: s.vanvoorst@fm.ru.nl

supranational executive actor: the European Commission. The Commission has a number of crucial tasks related to European legislation. Firstly, it is responsible for the development and formulation of legislative proposals (Schmidt and Wonka, 2013: 2). Secondly, it produces delegated and implementing acts (McCormick, 2015: 169–172). Thirdly, in its role of ‘guardian of the European treaties’, the Commission is responsible for monitoring and enforcing national compliance with European legislation (McCormick, 2015: 169–172; Schmidt and Wonka, 2013: 2).

These three tasks of the Commission in the EU’s legislative process have received ample academic scrutiny (e.g. Kassim et al., 2013; Schmidt and Wonka, 2013; Wille, 2013). Conversely, the literature has hardly touched upon a fourth key task of the Commission, which is to conduct *ex-post* evaluations that assess the functioning and effectiveness of European legislation. So far, such *ex-post* legislative (EPL) evaluations have mostly been neglected by scholars (but see Fitzpatrick, 2012; Mastenbroek et al., 2016; Zwaan et al., 2016), which is all the more surprising given both their theoretical importance and their growing role in the Commission.

Theoretically speaking, EPL evaluations may fulfil two important functions in legislative processes. Firstly, by recommending how the implementation of legislation can be improved and/or how legislation can be amended to increase its effectiveness, EPL evaluations are a potential tool for decision-makers to improve their policies (Fitzpatrick, 2012: 479; Vedung, 1997: 109). Secondly, EPL evaluations can be used to judge the performance of the actors that implement legislation, thus holding them accountable for their actions (Coglianese, 2012: 11; Vedung, 1997: 102–108).

Over the years the Commission has increasingly recognized the importance of EPL evaluations. It first emphasized the role of such evaluations in legislative improvement and accountability relationships in 2000, after which it started to make its procedures for EPL evaluations more systematic in 2007 (European Commission, 2007: 3–4; Fitzpatrick, 2012: 478). Since 2010 the Commission has also stressed the importance of EPL evaluations for judging the suitability of entire regulatory frameworks (so-called ‘fitness checks’) (European Commission, 2010: 5). Furthermore, from 2012 onwards it has given EPL evaluations a central place in its REFIT programme, which aims to identify and remove superfluous rules (European Commission, 2012: 4).

In 2015 the Commission published new guidelines that outline the methods, follow-up procedures and institutional responsibilities for carrying out EPL evaluations (European Commission, 2015). In principle, all the Commission’s EPL evaluations must use some form of stakeholder consultation to map the views of those actors that are directly affected by European legislation (European Commission, 2015: 299–336). Aside from this, EPL evaluations can use different combinations of methods, such as expert interviews, document analysis and quantitative modelling (European Commission, 2015: 337–414; Fitzpatrick, 2012: 490–497).

Concerning the follow-up of EPL evaluations, the Commission is supposed to produce an action plan based on the main recommendations of each evaluation to

ensure that its results feed back into the ‘regulatory cycle’ (European Commission, 2015: 297–298). Existing research has shown that the extent to which this happens varies in practice, such that about half of the ex-ante evaluations (impact assessments) attached to proposals for legislative amendments make use of information from EPL evaluations when available (van Golen and van Voorst, 2016: 388).

Concerning the institutional responsibility for EPL evaluations, the Commission’s guidelines specify that such evaluations are the responsibility of the Directorates-General (DGs), with a coordinating role for the Commission’s Secretariat-General (SG) (European Commission, 2015: 257; Stern, 2009: 70–71). EPL evaluations are usually based on reports written by external consultants to enhance their independence, but when this is more practical the whole evaluation process may also be conducted internally (European Commission, 2015: 282–289).

Importantly, since 2007 the Commission’s guidelines also prescribe that both financial and legislative activities must be evaluated periodically, in proportion to their allocated resources and expected impact (European Commission, 2007: 22; 2015: 257). In reality, however, not all important EU legislation is evaluated. Academic research has shown an initiation ratio of 33% for major EU regulations and directives from the period 2000 to 2012 (Mastenbroek et al., 2016: 1338). The European Commission (2013: 13) has produced similar figures: in 2013, 29% of all important EU regulations had been evaluated, with a further 13% of these regulations being evaluated at that moment, 19% of these regulations having a future evaluation planned and no numbers being provided for directives.

These figures show that the Commission is apparently selective in which legislation it evaluates, for reasons that the institution itself does not explain. This finding is problematic because an evaluation system is only credible if its procedures for initiating evaluations are systematic and transparent (Organisation for Economic Co-operation and Development (OECD), 2015: 120). If this is not the case, legislative quality may diminish in policy areas that are evaluated less frequently (OECD, 2015: 120) and/or the image could arise that the Commission decides what legislation to evaluate based on political considerations (Radaelli and Meuwese, 2010: 146). This, in turn, could harm the credibility of evaluations in the eyes of the legislator and other actors (Poptcheva, 2013: 4).

Therefore, this article looks into the question of what drives the initiation of EPL evaluations by the Commission. In other words, why does the Commission evaluate some pieces of law while it does not evaluate others? By answering this question, we not only seek to shed light on the unexplored topic of EPL evaluations in the EU, but also aim to further explore the motives that drive the Commission’s behaviour (Boswell, 2008: 472; Franchino, 2007: 11; Hartlapp et al., 2015: 1; Radaelli, 1999: 760–762; Wille, 2010: 1098–1100).

Two potential motives (not) to initiate an evaluation are studied in this paper, each of which is linked to a specific theoretical image of the Commission. The first motive, which is in line with the image of the Commission as the ‘guardian of the European treaties’, is the effective enforcement of EU legislation. Since EPL evaluations are a potential tool to check how legislation is implemented by the member

states (European Commission, 2015: 296; Stame, 2008: 124), we can expect that legislation for which the chances of non-compliance are higher is more likely to be evaluated. The second motive is the strategic protection of competences, which is in line with the image of the Commission as a political actor (Boswell, 2008: 472; Hartlapp et al., 2015: 1; Majone, 2005: 65). Following this logic, we would expect that the Commission refrains from evaluating legislation if this could result in a reduction of its powers.

The hypotheses flowing from these two motives are tested with the help of two datasets, the first containing all major EU legislation from 2000 to 2004 and the second containing all EPL evaluations conducted by the Commission during 2000–2014. With these data, we are able to draw conclusions about the Commission's decisions to evaluate European legislation over a 15-year period. The 10-year gap between our datasets is needed to give the Commission enough time to evaluate, thus avoiding any bias in our data in favour of legislation that was evaluated sooner. Binary logistic regression was used for the analysis.

Our results show that EU legislation is more likely to be evaluated if it is a directive rather than a regulation and if it is more complex, which is in line with the enforcement motive. Both of our control variables – the presence of evaluation clauses and the amount of evaluation capacity of the DG to which a piece of law belongs – also provide significant explanations. However, we did not find evidence that the strategic protection of competences explains the Commission's initiation of EPL evaluations.

Theoretical framework

Whereas evaluation-related topics are frequently discussed in the academic literature, there is no comprehensive approach to explaining why organizations decide to evaluate or not (Mastenbroek et al., 2016: 1343; Pattyn, 2014: 351). Therefore, this article develops such an approach in the context of the EU, building on two potential motives for the Commission: an enforcement motive and a strategic motive. These motives are closely linked to ongoing academic debates about the nature of the Commission (Boswell, 2008: 472; Franchino, 2007: 11; Hartlapp et al., 2015: 1; Radaelli, 1999: 760–762; Wille, 2010: 1098).

Enforcement motive

In its role of 'guardian of the European treaties', the Commission has the task to monitor and enforce member state compliance with EU legislation (Schmidt and Wonka, 2013: 2). EPL evaluations are potentially useful for this purpose, as they can collect and present information about how rules are implemented in practice (Coglianese, 2012: 11). This, in turn, makes EPL evaluations useful to hold those actors responsible for the implementation of legislation accountable (Vedung, 1997: 102). Therefore, EPL evaluations are a potential tool for the Commission to detect non-compliance by the member states and to address such

non-compliance via enforcement measures (European Commission, 2015: 292; Stame, 2008: 124).

The role of EPL evaluations in enforcing European legislation is also evident from earlier research about this topic. Mastenbroek et al. (2016: 1339) found that out of 216 EPL evaluations conducted or outsourced by the Commission between 2000 and 2012, 79% assessed the processes of legislative implementation, enforcement and/or compliance. Zhelyazkova et al. (2016: 833) found EPL evaluations to be the most detailed source of information about the compliance of member states with 24 directives of interest.

Those EPL evaluations that study member state compliance often assess the legal implementation of directives by systematically comparing national transposition measures, while they tend to assess the practical implementation of European legislation via surveys and interviews among stakeholders. In some cases, infringement data are also used as a source (Smith, 2015: 92–93). EPL evaluations that address member state compliance also tend to include recommendations for the Commission. Often these recommendations focus on ‘soft’ measures like increased monitoring, sharing best practices or publishing guidelines for national implementing authorities, but evaluations may also recommend the Commission to launch infringement procedures (Mastenbroek et al., 2016).

If the Commission can use EPL evaluations for enforcement purposes, we can expect that the chances that an evaluation is initiated are higher for pieces of law where there is a greater need to scrutinize the member states. In other words, we can expect that the chances that an evaluation is initiated are higher for legislation that offers more opportunities for non-compliance.

Three specific variables may be important in this regard. Firstly, the *type of legislation* may affect the chances of non-compliance. Directives offer the member states more discretion than regulations because they need to be transposed into national law (Treib, 2014: 6). In turn, this discretion offers the member states more opportunities to delay or prevent implementation (Kaeding, 2006: 232; König and Mäder, 2014: 247; Mastenbroek, 2003: 372; Steunenberg and Rhinard, 2010: 495; Treib, 2014: 6). Therefore, we expect directives to be more likely to be evaluated than regulations (Stame, 2008: 124).

H1: Directives are more likely to be evaluated than regulations.

Secondly, the *complexity of legislation* can affect the chances of non-compliance. Since the European legislative process includes multiple veto players – notably the Commission, the Council and the EP – decision making often produces compromises that are laid down in long and ambiguous texts (Häge, 2007: 307–308; Hofmann, 2013: 99). Such complexity offers member states more leeway for interpretation, and, therefore, makes it more difficult to establish whether they are complying with legislation or not (Kaeding, 2006: 242; König and Mäder, 2014: 253–254; Mastenbroek, 2003: 376; Steunenberg and Rhinard, 2010: 501). This in

turn can be expected to increase the chance that legislation is evaluated by the Commission.

H2: The more complex a piece of law, the higher its chances of being evaluated.

Thirdly, the political sensitivity of legislation may affect the chances of non-compliance. The more controversial a regulation or directive, the more likely it is that some member states who opposed it during the legislative process will not implement it correctly (Mastenbroek, 2003: 376). Since the Council represents the member states, *politicization in the Council* is especially likely to increase the chances of non-compliance (Treib, 2014: 14), and, therefore, the chances that an evaluation is initiated.

H3: The more politicized a piece of law was in the Council, the more likely it is to be evaluated.

Strategic motive

The hypotheses presented above are in line with the image of the Commission as the ‘guardian of the European treaties’. However, in recent years scholars have increasingly viewed the Commission as an actor that not only fulfils the tasks that the member states have delegated to it (such as enforcing European legislation), but also strategically pursues its own preferences (Franchino, 2007: 11; Hartlapp et al., 2015: 1–14; Wille, 2010: 1099). According to this political view on the Commission the institution has a perpetual interest to protect its competences, as without these competences it would not be able to achieve any of its (temporary) political aims (Hartlapp et al., 2015: 1; Majone, 2005: 65; Pollack, 2008: 9).

The Commission has been shown to deal strategically with ex-ante evaluations of legislation (impact assessments) (Poptcheva, 2013: 4; Torriti, 2010: 1065) and expert knowledge in general (Boswell, 2008: 472), so we can expect strategic considerations to play a role in decisions about the initiation of EPL evaluations as well. Ex-post evaluations are not just neutral instruments that can be used to stay informed about policy implementation, but also potential strategic tools that can strengthen or weaken the positions of actors (Bovens et al., 2008: 320; Schwartz, 1998: 295; Vedung, 1997: 111). As evaluations suggest changes to existing arrangements, they are inherently advantageous to some actors and disadvantageous to others (Bovens et al., 2008: 320; Weiss, 1993: 95–98). Negative evaluations can be particularly disadvantageous to actors that are responsible for delivering policies, as such evaluations may lead to demands to roll back their competences or to put them under closer supervision (Vedung, 1997: 102–108). This, in turn, may be an incentive for such actors to avoid evaluations that may have negative consequences (Schwartz, 1998: 295; Weiss, 1993: 95).

Therefore, we can expect the Commission to be reluctant to initiate EPL evaluations in situations where the results of such evaluations could be harmful to its interests. The Commission's better regulation agenda officially endorses the idea that EU legislation should be significantly amended or even repealed if an evaluation shows that it has no added value (European Commission, 2012: 3; 2013: 1; 2015: 254). In reality, however, we can expect that the Commission wants to avoid such situations to protect its competences (Majone, 2005: 65). In other words, we expect the chances that a piece of law is evaluated to be lower if the potential evaluation is more likely to be used to argue for significant amendments to the law.

Involvement of the European Parliament (EP) in decision making decreases the chances of significant amendments and is therefore expected to increase the chances that an evaluation is initiated. The reason for this is that the EP provides an extra veto player that can block amendments (Häge, 2007: 307; Hofmann, 2013: 102). As a majority of EP members generally supports further European integration (Pollack, 2008: 9), it can also be expected that the EP will usually oppose reducing the competences of supranational institutions like the Commission.

H4: Pieces of law that can only be amended with the approval of the European Parliament are more likely to be evaluated than pieces of law that can be amended without the approval of the European Parliament.

The *voting procedure in the Council* is also expected to influence the chances of legislative amendments. If unanimity is required in the Council it is significantly harder to change legislation, as it is difficult to make all member states agree on a proposal (Häge, 2007: 308). We therefore expect the chances that an evaluation is initiated to be higher when the Council applies unanimity voting, as compared to when it applies qualified majority voting (QMV).

H5: Pieces of law decided upon by unanimity in the Council are more likely to be evaluated than pieces of law decided upon by qualified majority voting.

Control variables

Aside from the two theoretical explanations described above, this research controls for two other potential explanations for decisions to initiate EPL evaluations. The first control variable is the *presence of an evaluation clause*. Many EU regulations and directives contain provisions that oblige the Commission to evaluate them after a number of years, which are usually inserted by the Council and the EP to ensure that they will stay informed about the legislation (Summa and Toulemonde, 2002: 410). We can expect legislation containing an evaluation clause to be evaluated more often than legislation without such a clause.

The second control variable is the *evaluation capacity of the responsible DG*. In this context, evaluation capacity is defined as the presence of sufficient means and

processes to ensure that evaluation is an ongoing practice in an organization (Nielsen et al., 2011: 325). Evaluation capacity includes the presence of organizational structures and procedures that support evaluations, the presence of sufficient financial and human capital to evaluate and the presence of proper (methodological) tools to conduct evaluations (Nielsen et al., 2011: 326–327). Since evaluation capacity varies primarily between the DGs of the Commission (Van Voorst, 2017: 25), we expect that legislation under the responsibility of DGs with higher evaluation capacity is more likely to be evaluated than other legislation.

We could also expect the Commission to be more likely to initiate an EPL evaluation if an ex-ante evaluation (impact assessment) of the same legislation was carried out, as impact assessments often contain a section prescribing that legislation should be evaluated ex-post (European Commission, 2015: 246–251). However, this variable cannot be studied in this research because the Commission's system for impact assessments was only set up in 2002–2003 (European Commission, 2007: 4).

Methods and data

Data collection

Although multiple datasets of EU legislation already exist (e.g. Hofmann, 2013: 102; Treib, 2014: 27), none of them suited the specific aims of our research. Therefore, we created two datasets for the task at hand, one containing major European legislation and one containing EPL evaluations (also see Mastenbroek et al., 2016).

The dataset of legislation covers the years 2000–2004. This period was chosen to give the Commission sufficient time to evaluate. While academic literature indicates that legislation is usually evaluated after about five years (Eijlander and Voermans, 2000: 355) and evaluation clauses in EU legislation also tend to give the Commission five years or less to evaluate, we decided to double this period to avoid concluding that any legislation has not been evaluated while an evaluation was in fact still upcoming. Therefore, our dataset of legislation stops at the end of 2004, but it should be emphasized that our article concerns decisions to initiate EPL evaluations over a period of 15 years (2000–2014), which is further explained by the description of our second dataset below.

Because the Commission follows the logic that the resources spent on an evaluation must be proportionate to the importance of a measure (European Commission, 2007: 22; 2015: 255–256), minor EU legislation does not have to be evaluated (European Commission, 2015: 253). Therefore, our dataset of legislation only includes major regulations and directives. We excluded all delegated and implementing acts,¹ which are generally considered less important than primary legislation (Franchino, 2007: 80), as well as all rectifications, amendments and secondary Council legislation. Because of the explicit link between evaluations and improving the effects of legislation on European citizens and companies (European Commission, 2007: 3; 2012: 2; 2013: 1–2), we also excluded legislation without direct relevance for national actors. This includes legislation that only

addresses EU institutions or foreign countries. Together, the selection criteria led to a dataset of 277 major directives and regulations adopted in the period 2000–2004. Our dataset of evaluations (see below) contains only eight evaluations of legislation that we did not consider ‘major’ (2% of all evaluations), indicating that our selection criteria were fairly appropriate.

To assess the initiation ratio of evaluations, our dependent variable, we extracted information from a second dataset. This dataset contains 313 EPL evaluations of regulations, directives and treaty articles conducted or outsourced by the Commission between 2000 and 2014 (Mastenbroek et al., 2016: 1334–1335).² Evaluations completed before 2000 were omitted because of a lack of data, and evaluations merely containing studying prescriptions for foreign countries and EU institutions were excluded for the same reasons as discussed above. We also discarded those evaluation reports that merely summarize other evaluations.

The evaluations were gathered from different sources: The Commission’s multi-annual evaluation overview (2010), the Commission’s search engine for evaluations,³ the Commission’s work programmes,⁴ EU bookshop,⁵ annexes to the Commission’s financial reports,⁶ and lists of evaluations found on the websites of DGs. We checked our data using an existing list of evaluations produced by expertise centre Eureval, by running Google searches for evaluations of all major legislation adopted between 1996 and 2010, by searching for background documents of legislation in Eur-lex,⁷ and by discussing our data-gathering method with the SG (for a further description of the dataset of EPL evaluations, see Mastenbroek et al., 2016: 1334–1335).

Operationalization

Starting with the enforcement motive, the *type of legislation* (hypothesis 1) was measured as a dichotomous variable (directive or regulation). The *complexity of legislation* (hypothesis 2) was measured through its number of recitals, as more complex legislation generally requires a larger number of explanations (Franchino, 2000: 74; Kaeding, 2006: 236; Steunenberg and Rhinard, 2010: 501; Treib, 2014: 26). *Politicization in the Council* (hypothesis 3) was measured by determining if a legislative proposal was on the Council’s agenda as a *B-point*, as B-points represent the topics that are actively debated at the political level (Häge, 2007: 303; Hofmann, 2013: 126; König, 2008: 149).

Concerning the strategic motive, *involvement of the European Parliament* (hypothesis 4) was measured by looking at the formal procedure used to enact the legislation as stated by Eur-lex. In case of the ordinary legislative procedure (former codecision and cooperation procedures) this involvement was considered high, while in case of the consultation procedure it was considered low (Häge, 2007: 316). The *voting procedure in the Council* (hypothesis 5) was also measured as a dichotomous variable (QMV or unanimity) using Eur-lex.

Concerning the control variables, we searched each piece of law using specific keywords⁸ to establish the *presence of an evaluation clause* (yes/no). We also

checked the last five articles of each regulation or directive, as this is the most common place for evaluation clauses. Concerning evaluation capacity, we measured 12 indicators derived from a model developed by Nielsen et al. (2011, 326–330) via interviews with the European Commission (Van Voorst, 2017: 29–31). However, only the presence of a specialized evaluation (sub-)unit (yes/no) and the presence of evaluation guidelines (yes/no) could be established per DG per year and were, therefore, useful as indicators. For legislation that has been evaluated, the data used concern the year when the evaluation was published. For legislation that has not been evaluated, we assumed that this decision was made five years after the legislation was published (the modal value of the time between publication dates of legislation and publication dates of evaluations is six years in our dataset, from which we subtracted one year as evaluations usually take that long to conduct) and determined the scores for evaluation capacity accordingly. However, because this assumption of five years is somewhat arbitrary we also experimented with other time periods, which affected our results to some extent.⁹

The operationalization of all our variables is summarized in the table in the Online appendix. Because of the binary nature of our dependent variable, logistic regression was used for the analysis. The variables belonging to the two motives to evaluate were entered as blocs to allow for comparisons between the models.

Results

Out of the 277 major regulations and directives in our dataset, 116 have been evaluated ex-post. This is an initiation ratio of 41.9%, meaning that about six out of 10 major pieces of EU law from 2000 to 2004 have not (yet) been evaluated by the Commission. This initiation ratio is higher than the 33% found during earlier research about major legislation from 2000 to 2002 (Mastenbroek et al., 2016: 1338), indicating that legislation published during 2003–2004 was evaluated more often than older legislation. This is a sign that the proportion of legislation evaluated by the Commission may be increasing over time.

A few pieces of law in our dataset were evaluated multiple times over the years: 15 pieces of law were evaluated twice, four pieces of law were evaluated thrice and two pieces of law were evaluated four times. Due to the binary nature of our dependent variable, these pieces of law with more than one evaluation have no special impact on our analysis: they were simply coded as 1. Their number was also too low to conduct an additional analysis of the number of times that a piece of law was evaluated.

Figure 1 below depicts the initiation ratio per DG. The three DGs with the highest initiation ratios are DG Eurostat (71.4%), DG Competition (66.0%) and DG Internal Market (65.2%). DG Trade and DG Economic and Financial Affairs have not evaluated their few major pieces of law from 2000 to 2004 at all; besides this the three DGs with the lowest initiation ratios are DG Energy (28.6%), DG Home Affairs (28.0%), and DG Agriculture (20.8%). The variation among DGs is included in the analysis through the evaluation capacity variables; the data do not

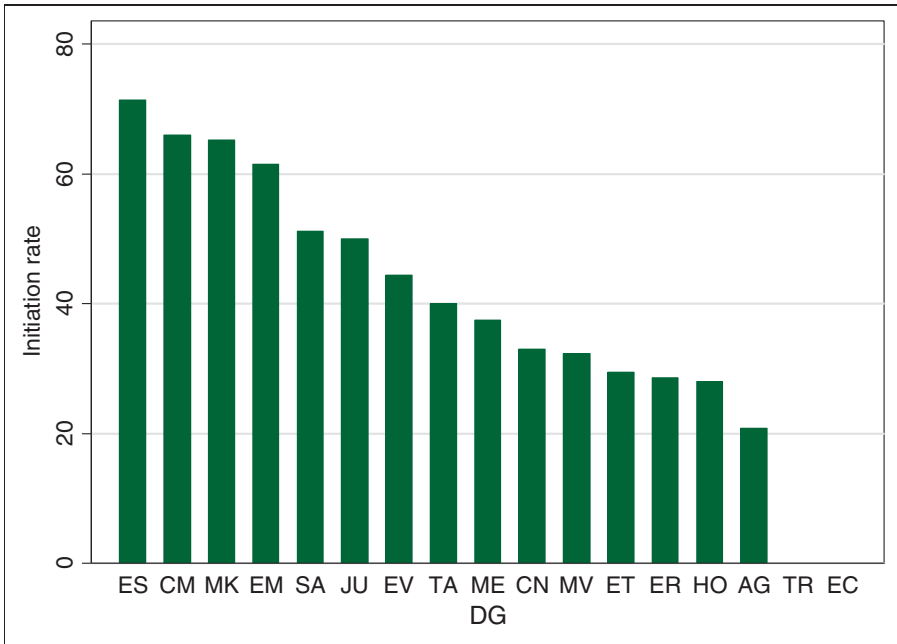


Figure 1. Initiation ratio per DG.

Note: AG: agriculture; CM: competition; CN: communications and technology; DG: Directorates-General; EC: economic and financial affairs; EM: employment; ER: energy; ES: Eurostat; ET: enterprise and industry; EV: environment; HO: home affairs; JU: justice; ME: maritime affairs; MK: internal market; MV: transport; SA: health and consumers; TA: taxation; TR: trade. Some DGs have merged and/or changed their names since 2014.

suggest other patterns concerning the size or policy areas of the DGs that warrant investigation.

Table 1 below presents the explanatory analysis. The results show that the model with the variables belonging to the enforcement motive and the control variables only passes the chi-square test and is therefore significant. Aside from politicization in the Council all the individual variables included in this model are also significant. Table 1 furthermore shows that if the variables related to the strategic motive are added, the model as a whole and the individual variables that were significant before are still significant. However, neither the involvement of the EP nor the voting procedure in the Council provides an explanation for variation in the initiation of EPL evaluations. Below each variable will be discussed in detail.

Starting with the *type of legislation*, the first variable belonging to the enforcement model, Table 1 shows that the odds of being evaluated are about 2.05 times higher for directives than for regulations. In terms of predicted probabilities – which are easier to interpret than odds ratios – the chances of an evaluation taking place are 14.0% higher for directives than for regulations, if all other variables are kept at their observed values. In terms of descriptive statistics, out of the

Table 1. Results of the logistic regression.

	Model 1: Enforcement motive			Model 2: Strategic motive		
	B (SE)	Sig.	Odds ratio	B (SE)	Sig.	Odds ratio
Constant	-2.57 (0.41)	.00	0.08	-2.65 (0.43)	.00	0.07
Legislation type	0.66 (0.29)	.02	1.93	0.72 (0.30)	.02	2.05
Complexity	0.03 (0.01)	.02	1.03	0.03 (0.01)	.02	1.03
Politicization	-0.51 (0.32)	.11	0.60	-0.50 (0.32)	.12	0.60
Council voting				-0.15 (0.36)	.69	0.86
EP involvement				0.46 (0.49)	.35	1.58
Clause	1.57 (0.33)	.00	4.79	1.55 (0.34)	.00	4.69
Unit	0.87 (0.31)	.01	2.38	0.95 (0.32)	.00	2.59
Guidelines	0.73 (0.30)	.01	2.08	0.82 (0.31)	.01	2.26
<hr/>						
N	277			277		
Chi ²	62.70			64.62		
Sig	0.00			0.00		
McFadden R ²	0.17			0.17		
AIC	1.184			1.191		

AIC: Akaike information criterion.

141 major directives in our dataset, 51.8% have been evaluated; out of the 136 major regulations, only 31.6% have been evaluated. In line with hypothesis 1, these findings indicate that the Commission prioritizes evaluating directives over regulations.

The *complexity of legislation*, the second variable belonging to the enforcement model, also significantly increases the chances of an evaluation occurring. For every extra recital, the odds of a piece of law being evaluated increase by about 3%. Figure 2 below presents the effect of this variable in terms of predicted probabilities. The figure shows that the chances of an evaluation occurring increase from about 0.3 to 0.7 as the number of recitals grows, with an average growth in predicted probability of 0.06% per recital, if all other variables are kept at their observed values. These findings are in line with hypothesis 2.

Politicization in the Council, the third variable belonging to the enforcement model that was measured by the occurrence of the legislative proposal as a B-point on the Council’s agenda, is not significant. Accordingly, we reject hypothesis 3.

Turning to the political variables, the results in Table 1 show that neither the involvement of the EP nor the voting procedure in the Council provides a significant explanation for variation in the initiation of EPL evaluations. This means that hypotheses 4 and 5 are rejected. These results indicate that the chances of a piece of

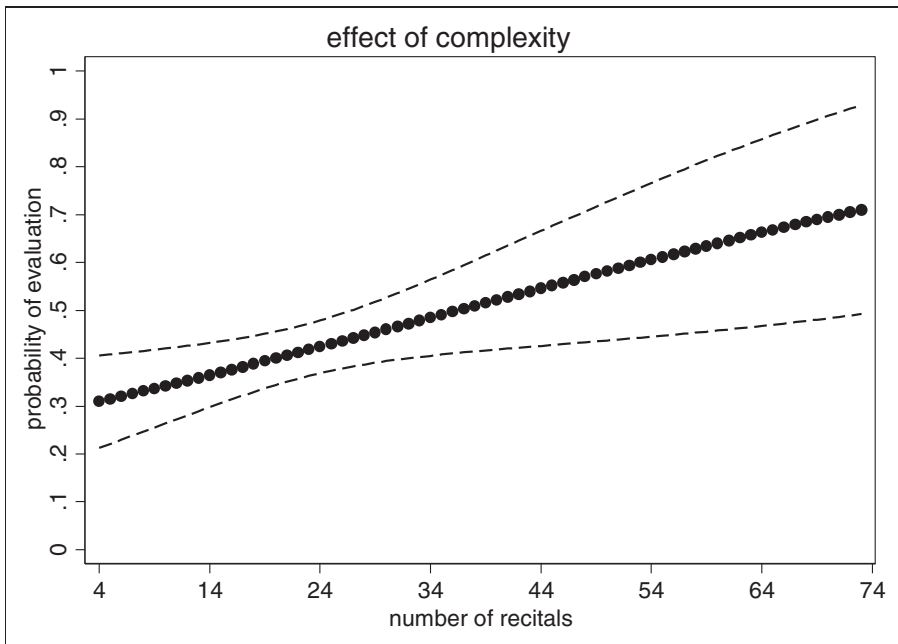


Figure 2. Effect of legislative complexity on the probability of an evaluation occurring.

law being significantly amended do not affect the Commission's decision to evaluate it or not.

Conversely, both control variables turn out to be significant. Table 1 shows that the odds of a piece of law being evaluated become about 4.69 times higher if an evaluation clause is present as compared to legislation without such a clause. In terms of predicted probabilities, the chances of an evaluation taking place are 30.9% higher for legislation with an evaluation clause than for legislation without such a clause, if all other variables are kept at their observed values.

Despite the significance of this variable, it should be noted that only 92 out of 165 pieces of law with an evaluation clause (55.8%) were evaluated, while 24 out of 112 pieces of law without such a clause (21.4%) were evaluated as well. The first number shows that the Commission only complied with a little more than half of the evaluation clauses inserted in major legislation from 2000 to 2004, indicating that the presence of such clauses is not a guarantee that legislation will be evaluated. The numbers also show that the presence of evaluation clauses only explains a part of the variation in the initiation of EPL evaluations.

Table 1 also shows that both indicators for evaluation capacity are significant. The odds of a piece of law being evaluated are about 2.59 times higher for legislation of a DG with an evaluation unit as compared to legislation of a DG without such a unit, and 2.26 times higher for legislation of a DG that has evaluation guidelines as compared to legislation of a DG without such guidelines. In terms

of predicted probabilities, the chances of an evaluation taking place are 18.3% higher in the first case and 15.9% higher in the second case, if all other variables are kept at their observed values.

When interpreting these results, however, it should be noted that high evaluation capacity may be a consequence as well as a cause of evaluation-related activities. For example, it is possible that DGs that initiate more EPL evaluations also invest more in evaluation guidelines to support such evaluative activities. It should also be noted that the results concerning evaluation capacity somewhat depend on our assumptions about the number of years after which it was decided not to evaluate certain legislation (as explained in our methodology section and Note 9). Therefore, more research is needed to establish the exact effect of evaluation capacity on the initiation of EPL evaluations in the EU.

Conclusion

This article has sought to describe and explain the variance in the initiation of EPL evaluations by the European Commission. Although the Commission officially endorses EPL evaluations (European Commission, 2007: 3; 2013: 11; 2015: 296), little was known about how systematically the institution conducts such evaluations in practice (but see Mastenbroek et al., 2016). This study aimed to shed light on this underexplored topic by developing a theoretical approach based on two motives for the Commission to evaluate – an enforcement and a strategic motive – while controlling for other potential explanations. We tested these explanations with the help of binary logistic regression, based on two self-developed datasets.

The results show that less than half of all major EU legislation from 2000 to 2004 (41.9%) was evaluated. However, the proportion of evaluated legislation has increased over time. Only a small proportion of the major legislation was evaluated more than once.

Concerning the enforcement motive, our results suggest that the odds of being evaluated are significantly higher for directives than for regulations, and that these odds also increase significantly as legislation becomes more complex. This indicates that the Commission prioritizes evaluating legislation for which the chances of non-compliance are relatively high, and that evaluations may at least partly be initiated to scrutinize member state implementation. Concerning the strategic motive, however, we did not find any significant results. This indicates that the risk of EU legislation being significantly amended does not affect its odds of being evaluated.

Two control variables also turned out to be significant. Firstly, the odds of legislation being evaluated increase significantly if that legislation contains an evaluation clause. However, our data also revealed that the Commission only complies with such clauses in about half of all cases. Secondly, the evaluation capacity of the DG that is responsible for the legislation significantly increases the odds of that legislation being evaluated.

In conclusion, our analysis indicates that the initiation of EPL evaluations by the Commission is best explained by a mix of its need to enforce EU legislation towards the member states, its formal obligations to evaluate and its evaluation capacity. However, these conclusions should be viewed in the light of two possible limitations of this research. Firstly, the quantitative nature of our study required us to use indicators that could be measured efficiently for a large number of cases. Some of these indicators may not entirely cover the abstract concepts that they are supposed to represent, such as evaluation capacity and politicization. Therefore, to sustain the conclusions of this article, a follow-up case study with more sophisticated indicators would be useful.

A second limitation of this study is its time period. As explained above, EPL evaluations may be conducted a decade or more after a piece of law enters into force, so we could not yet assess the extent to which legislation from after 2004 has been evaluated without risking a bias in our data. Whereas our dataset of 277 major regulations and directives (initiated by seventeen DGs) is so broad that our findings are probably not affected by any particular political choice made during 2000 to 2004, it still seems worthwhile to repeat this research in the future to assess to what extent post-2005 legislation is evaluated.

Two other possibilities for future research stand out. Firstly, since this article showed that the Commission does not always comply with evaluation clauses, a follow-up study about the reasons for this seems worthwhile. Secondly, it could be examined to what extent the factors presented in this study also explain variance in the quality of EPL evaluations, as this is another important characteristic of a proper evaluation system and previous research has shown that the quality of the Commission's EPL evaluations varies greatly (Mastenbroek et al., 2016: 1340–1341).

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Supplemental material

Supplementary material is available for this article online.

Notes

1. This refers to all legislation having its legal basis in another regulation or directive.
2. Evaluations published in 2013 and 2014 were added to the dataset used in Mastenbroek et al. (2016), resulting in a total of 313 cases. Six further evaluations published in French were excluded because of our lacking language skills. However, including these evaluations would not have changed the results, since they did not study additional legislation from 2000 to 2004.
3. ec.europa.eu/smart-regulation/evaluation/search/search.do
4. http://ec.europa.eu/atwork/key-documents/index_en.htm
5. <https://bookshop.europa.eu/en/home/>

6. SWD(2013)228 and SWD(2012)383.
7. <http://eur-lex.europa.eu/homepage.html>
8. 'evalu*', 'repo*', 'stud*' and 'research'.
9. As described in the methodology section of our article, for legislation that was not evaluated we entered our evaluation capacity data based on the assumption that decisions not to evaluate were taken five years after the publication date of the legislation. We experimented with other time periods than five years for this as well (four, six, seven and eight years). If we assume decisions not to evaluate to be taken after six years instead of five the presence of an evaluation unit is no longer significant, and if we assume this period to be seven years or more both indicators for evaluation capacity cease to be significant. However, assuming a lesser number of years than five does not change the significance of either indicator. The reason for these different findings appears to be that some of the earliest DGs to develop evaluation units and/or guidelines (DG Eurostat, DG Internal Market and DG Employment in particular) also have the highest initiation ratios (as listed in Figure 1). Around 2010 many DGs that have smaller initiation ratios also started to develop such evaluation capacity. Therefore, belonging to a DG that was a frontrunner in building evaluation capacity could be more important than belonging to a DG with high evaluation capacity per se when it comes to explaining variation in the initiation of EPL evaluations.

References

- Boswell C (2008) The political functions of expert knowledge: Knowledge and legitimation in the European Union. *Journal of European Public Policy* 15(4): 471–488.
- Bovens M, 't Hart P and Kuipers S (2008) The politics of policy evaluation. In: Goodin RE, Rein M and Moran M (eds) *The Oxford Handbook of Public Policy*. Oxford: University Press, pp. 320–335.
- Coglianesi C (2012) Evaluating the performance of regulation and regulatory policy. Report to the OECD.
- Eijlander P and Voermans W (2000) *Wetgevingsleer [Legislative theory]*. The Hague: Boom Juridische Uitgevers.
- European Commission (2007) *Communication to the Commission from Ms Grybauskaitė in Agreement with the President: Responding to Strategic Needs: Reinforcing the Use of Evaluation [SEC(2007)213]*. Brussels: European Commission.
- European Commission (2010) Multi-annual overview (2002–2009) of evaluations and impact assessments. Available at: http://ec.europa.eu/smart-regulation/evaluation/docs/multiannual_overview_en.pdf (accessed July 2015).
- European Commission (2012) *EU Regulatory Fitness [COM(2012)746]*. Brussels: European Commission.
- European Commission (2013) *Regulatory Fitness and Performance (REFIT): Results and Next Steps [COM(2013)685 final]*. Brussels: European Commission.
- European Commission (2015) *Better Regulation Toolbox [SWD(2015)111]*. Brussels: European Commission.
- Fitzpatrick T (2012) Evaluating legislation: An alternative approach for evaluating EU internal market and services law. *Evaluation* 18(4): 477–499.
- Franchino F (2000) Control of the Commission's executive functions. *European Union Politics* 1(1): 63–92.

- Franchino F (2007) *The Powers of the Union: Delegation in the EU*. Cambridge: University Press.
- Häge FM (2007) Committee decision-making in the Council of the European Union. *European Union Politics* 8(3): 299–328.
- Hartlapp M, Metz J and Rauh C (2015) *Which Policy for Europe? Power and Conflict Inside the European Commission*. Oxford: University Press.
- Hofmann A (2013) *Strategies of the repeat player. The European Commission between courtroom and legislator*. PhD Thesis, University of Cologne, Germany.
- Kaeding M (2006) Determinants of transposition delay in the European Union. *Journal of Public Policy* 26(3): 229–253.
- Kassim H, Peterson J, Bauer MW, et al. (2013) *The European Commission of the Twenty-First Century*. Oxford: University Press.
- König T (2008) Analysing the process of EU Legislative decision-making: To make a long story short. *European Union Politics* 9(1): 145–165.
- König T and Mäder L (2014) The strategic nature of compliance: An empirical evaluation of law implementation in the central monitoring system of the European Union. *American Journal of Political Science* 58(1): 246–263.
- McCormick J (2015) *European Union Politics*, 2nd ed. London: Palgrave.
- Majone G (1999) The regulatory state and its legitimacy problems. *West European Politics* 22(1): 1–24.
- Majone G (2005) *Dilemmas of European Integration: The Ambiguities and Pitfalls of Integration by Stealth*. Oxford: University Press.
- Mastenbroek E (2003) Surviving the deadline: The transposition of EU directives in the Netherlands. *European Union Politics* 4(4): 371–396.
- Mastenbroek E, Voorst S van and Meuwese A (2016) Closing the regulatory cycle? A meta-evaluation of ex-post legislative evaluations by the European Commission. *Journal of European Public Policy* 23(9): 1329–1348.
- Nielsen SB, Lemire S and Skov M (2011) Measuring evaluation capacity: Results and implications of a Danish study. *American Journal of Evaluation* 32(3): 324–344.
- Organisation for Economic Co-operation and Development (OECD) (2015) *OECD Regulatory Policy Outlook 2015*. Paris: OECD Press.
- Pattyn V (2014) Why organizations (do not) evaluate? Explaining evaluation activity through the lens of configurational comparative methods. *Evaluation* 20(3): 348–367.
- Pollack MA (2008) *Member-State Principals, Supranational Agents, and the EU Budgetary Process, 1970–2008*. Paper presented at the Conference on Public Finances in the European Union in Brussels, 3–4 April 2008.
- Poptcheva EM (2013) *Library briefing. Policy and legislative evaluation in the EU*. Brussels: European Parliament.
- Radaelli CM (1999) The public policy of the European Union: Whither politics of expertise? *Journal of European Public Policy* 6(5): 757–774.
- Radaelli CM and Meuwese ACM (2010) Hard questions, hard solutions: Proceduralisation through impact assessment in the EU. *West European Politics* 33(1): 136–153.
- Schmidt SK and Wonka A (2013) The European Commission. In: Jones E, Menon A and Weatherill S (eds) *The Oxford Handbook of the European Union*. Oxford: University Press, pp. 336–349.
- Schwartz R (1998) The politics of evaluation reconsidered: A comparative study of Israeli Programs. *Evaluation* 4(3): 294–309.

- Smith M (2015) Evaluation and the salience of infringement data. *European Journal of Risk Regulation* 6(1): 90–100.
- Stame N (2008) The European project, federalism and evaluation. *Evaluation* 14(2): 117–140.
- Stern E (2009) Evaluation policy in the European Union and its institutions. In: Trochim WMK, Mark MM and Cooksy LJ (eds) *Evaluation Policy and Evaluation Practice: New Directions for Evaluation*. San Francisco, CA: Jossey-Bass, pp. 67–85.
- Steunenberg B and Rhinard M (2010) The transposition of European Law in EU Member States: Between process and politics. *European Political Science Review* 2(3): 495–520.
- Summa H and Toulemonde J (2002) Evaluation in the European Union: Addressing complexity and ambiguity. In: Furubo J, Rist RC and Sandahl R (eds) *International Atlas of Evaluation*. New Brunswick: Transaction Publishers, pp. 407–424.
- Torriti J (2010) Impact assessment and the liberalization of the EU energy markets: Evidence-based policy-making or policy-based evidence-making? *Journal of Common Market Studies* 48(4): 1065–1081.
- Treib O (2014) Implementing and complying with EU governance outputs. *Living Reviews in European Governance* 9(1): 1–47.
- van Golen T and van Voorst S (2016) Towards a regulatory cycle? The use of evaluative information in impact assessments and ex-post evaluations in the European Union. *European Journal of Risk Regulation* 7(2): 388–403.
- Van Voorst S (2017) Evaluation capacity in the European Commission. *Evaluation* 23(1): 24–41.
- Vedung E (1997) *Public Policy and Program Evaluation*. New Brunswick: Transaction Publishers.
- Weiss CH (1993) Where politics and evaluation research meet. *American Journal of Evaluation* 14(1): 93–106.
- Wille A (2010) Political-bureaucratic accountability in the EU Commission: Modernising the executive. *West European Politics* 33(5): 1093–1116.
- Wille A (2013) *The Normalization of the European Commission: Politics and Bureaucracy in the EU Executive*. Oxford: University Press.
- Zhelyazkova A, Kaya C and Schrama R (2016) Decoupling practical and legal compliance: Analysis of member states' implementation of EU policy. *European Journal of Political Research* 55(4): 827–846.
- Zwaan P, Voorst S van and Mastebroek E (2016) Ex-post regulatory evaluations in the European Union: Questioning the use of evaluations as instruments for accountability. *International Review of Administrative Sciences* 82(4): 674–693.