

The role of agencification in achieving value-for-money in public spending

Luciana Cingolani¹ | Mihály Fazekas²

¹Hertie School of Governance

²Central European University

Abstract

Agencification has been pursued globally under the promise of increasing public administration performance. In spite of ample theoretical arguments, the empirical evidence on the causal link between agencification and performance remains scarce and methodologically contested. We contribute to this debate by empirically testing the impacts of agencification across Germany, Spain, and the United Kingdom on value-for-money, competitiveness, and timeliness during the period 2006–2016. We use unique administrative datasets, enabling objective and granular measurements of reforms and their effects, employing quasi-experimental methods. Findings suggest heterogeneous effects both across countries and outcomes. On average, value-for-money improves by 2.8% or 1.7 billion EUR over a decade, while outputs and processes change only marginally. Recently agencified organizations barely improve their performance, while older agencies achieve substantial improvements. The three countries' heterogeneous administrative contexts play a critical role as mediating factors, with the biggest changes occurring in higher new public management take-up countries.

1 | INTRODUCTION

Discussions on the promises and perils of creating semiautonomous single-purpose public bodies—also known as agencification—came near to turning four decades. Within the multiple streams of arguments

Luciana Cingolani and Mihály Fazekas contributed equally to this study.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Governance* published by Wiley Periodicals, Inc.

and findings coming from political proponents, opponents, and numerous research efforts, a wide set of issues around agencification have seen the light.

In the simplified terms, the idea of moving toward a higher bureaucratic autonomy model for certain policy areas has been championed under the theoretical expectation that greater managerial discretion, coupled with tighter center-of-government accountability and results control, would minimize moral hazard and adverse selection problems in the public sector. This dual change would turn policymaking more efficient and less politicized (e. g. Hood, 1991, Osborne & Gaebler, 1992). Such expectations have certainly not remained unchallenged, as reflections around the link between greater autonomization and performance have flourished in multiple directions, often questioning the resulting fragmentation of the public sector (e. g. Christensen & Læg Reid, 2007), or highlighting the necessary conditions for the expectations to be fulfilled. Similarly, a prolific discussion has taken place regarding the very definition of agencification and what exactly an agency is (Læg Reid & Verhoest, 2010; Verhoest, Peters, Bouckaert, & Vermeulen, 2004).

Yet, by multiple accounts, the slowest development in the field during the last 40 years has taken place regarding the empirical evidence on how agencification affects performance (Overman and van Thiel 2016; Dan, 2014; Pollitt & Dan, 2011). A comprehensive literature review on the effects of new public management (NPM) reforms in Europe finds that out of the 500+ studies under scrutiny—where a 14% dealt with the creation of semiautonomous units—only about one-fifth was conducted through empirical work (Dan, 2014; Pollitt & Dan, 2011). Of this empirical work, only a small fraction followed a carefully designed causal identification strategy capable of going beyond mere correlations. Moreover, it highlights that the operationalization of performance has mostly focused on capturing changes in internal processes and much less on outputs or outcomes. This overview, which includes works from both researchers and practitioners, paints a grim picture of the empirical foundations upon which dozens of countries around the world have undergone major restructuring reforms.

This article contributes to the empirical analysis of the impacts of agencification on the performance of public bodies. It does so by making a number of methodological improvements that complement and enhance existing efforts. First, it exclusively draws on hard administrative records describing organizational behavior in three different countries—Germany, Spain, and the United Kingdom—over an 11-year period, as opposed to using perception-based or self-reported behavioral data. Second, the study does not draw on a representative sample of observations but rather analyzes the full universe of agencies and their relevant transactions in selected countries. Turning the attention to the agency and transactional levels allows for performance measurement and impact-tracing mechanisms that better match the premises embedded in NPM theories. Third, it directly compares results in terms of procedural efficiency, immediate outputs resulting from those procedures, and more substantial outcomes in the form of value-for-money, an almost unexplored component of performance. Fourth, organizational performance is measured through a genuinely cross-cutting field: public contracting. Such a narrow focus enables the comparison of very different agencies in a consistent manner. Finally, we make use of a unique, large-scale dataset that enables us to employ quasi-experimental methods in order to better capture true casual effects and decompose results by country as well as outcome type.

Our findings depict a diverse picture with a variety of effects taking place: agencification improves a number of performance dimensions, while negatively affecting others. Most importantly, agencification appears to exercise a consistent positive impact on outcomes such as lowering prices paid by the public sector, while outputs and organizational processes may not change much. Our empirical results also confirm that the maturity of the agency (judged by the number of years in

place) exercises a decisive role in producing desirable outcomes, with older agencies being responsible for most of the positive impacts identified. Furthermore, important cross-national differences are observed, suggesting that the broader administrative context is critical in mediating agencification's effects, in line with prior theoretical knowledge. In this sense, the biggest changes are observed in the most thorough implementers of NPM reforms, namely the United Kingdom and Spain. Taken together, the impacts of agencification appear to be heterogeneous, context-dependent, and involve important nuances.

The next section outlines the main theoretical expectations around agencification and presents some of the most prominent empirical works aiming to put those expectations to test. It also offers some reflections on the reasons behind the slow progress on the empirical testing of the theoretical arguments. Section 3 provides a brief contextualization of agencification processes in each of our three country cases. Section 4 presents our hypotheses, empirical model, data, and indicators, while Section 5 reports the resulting findings and addresses the study's limitations. Section 6 resumes the discussion on the effects of agencification in light of our empirical results. The final section offers a few brief closing remarks.

2 | AGENCIFICATION: COMPACT THEORY AND INCONCLUSIVE EMPIRICS

The momentum behind the agencification wave that swept across multiple polities over the last decades can be explained at least in part by its theorized benefits, largely encompassed under the umbrella of the NPM paradigm. The creation of semiautonomous units most frequently consists in disaggregating ministerial units into smaller narrow-purpose organizations that are generally entrusted with executing ministerial or state policy. This way, the roles of policy planning and evaluation are separated from that of policy execution. Policy execution by semiautonomous entities would have the general advantage of providing more speedy responses and higher managerial efficiency, due to the involvement of highly specialized and invested managers, coupled with the support of various competency frameworks (Hood & Lodge, 2004). In addition, a higher autonomy from politics would provide regulated markets with greater predictability and credible commitment, reducing the chance of intertemporal inconsistencies (Kyland & Prescott, 1977). In exchange, agencies would respond to a tight system of performance management designed by the oversight ministries and be the subject of strong accountability (Pollitt et al 2001; Pollitt et al 2004). Against this backdrop, a variety of result-control mechanisms proliferated around newly created agencies (Verhoest, 2005), and the coordination capacity of central governments was put to the test. Agencification, moreover, bores the promise of improving both managerial processes and outcomes: in terms of processes, agencies would adapt faster to a changing environment, and in terms of outputs, they would organize all resources around goal achievement (Osborne & Gaebler, 1992).

As the knowledge around agencification evolved, these general principles were contested in light of the empirical reality in different countries and the actual difficulties of balancing autonomy and political control (Christensen & Lægheid, 2007). More nuanced discussions emerged, focusing, for example, on the difference made by the various and often contradictory types of autonomy (Verhoest et al., 2004; Yesilkagit, 2004), the alternative definitions of agency (Pollitt et al., 2004), or how the political economy of the affected polities ("local editing") worked (Sulle, 2010). In regard to the importance of context mediating the developments and achievements of agencification, the work of Moynihan (2006) posits that a single reform label normally masks an important variation in outcomes in response to policymakers' "*room for interpretation*" of such reforms. In this sense, a long

tradition in public policy studies acknowledges the influence of context in shaping reforms, including past legacies, political institutions, economic context, and social values, among others (see Moynihan, 2006 for a succinct overview). It is illustrative, for example, that a general agreement has been built around the idea that certain administrative traditions facilitate a more widespread take-up of NPM reforms (McLaughlin, Osborne, & Ferlie, 2002) and that more comprehensive commitments toward NPM tools increase the overall effectiveness of each individual tool; a pattern can be largely corroborated in cross-country studies (e.g., Hammerschmid, Van de Walle, Andrews, & Bezes, 2016, among others).

On the empirical front, however, the field of agencification is much less developed. Governments have collected some anecdotal evidence in order to conduct cost–benefit analyses of their reforms, while scholars have sought to show cross-country comparative results, albeit often facing important methodological limitations. Of the existing efforts, a majority of the efforts suggest modest evidence of a positive impact, yet many salient works only capture a link that is mediated by other specific factors or no link at all.

It is important to note that while some evidence exists on the role played by autonomization more generally (see Vining, Laurin, & Weimer, 2015), this condensed survey focuses exclusively on agencification. It does so by first reviewing the studies operating at the macrocomparative level and subsequently exploring other works that offer some measure of performance at the organizational level.

In an attempt to circumvent critiques on the paucity of outcomes as dependent variables, Overman and van Thiel (2016) seek to assess whether the number of existing agencies in 20 countries is able to explain organizational performance on different dimensions, including one understood as value-for-money. At this cross-country macrolevel, the study finds a statistically significant negative effect of agencification on performance in most of their models, which is also the case for value-for-money as an outcome. Value-for-money is here measured through an original index that divides a government performance score per country by the respective GDP percentage spent by the government.

Brewer (2004) contributes to the study of mediating factors by exploring the types of administrative reforms that improve bureaucratic performance in 25 Western democracies. It finds mixed results, with “flattening bureaucracy”—the closest fit to agencification—not achieving statistical significance. The study also suggests that contextual factors play a stronger role in determining results than the reforms themselves. In this sense, it lists a series of political risk factors that hinder bureaucratic performance: security threats, ethnic or religious tensions, divided government, and economic instability.

While not directly concerned with performance as an outcome, the work of Wynen and Verhoest (2016) unravels interesting angles in the link between autonomy and performance-related processes. In a comparative work involving nine countries (eight European and one Asian), the study tests the effect of organizational autonomy paired with external results control on the take-up of internal performance management in lower-hierarchy levels of public sector organizations. Their study uses agency-level survey data to executives from semiautonomous agencies, which share perceptions about *de facto* levels of autonomy and results control. Findings show that the introduction of both the result control and financial autonomy improves performance management take-up, while no effect is observed for personnel management autonomy. The same results are found in a smaller study, conducted earlier, of 226 semiautonomous agencies in Belgium, Ireland, and Norway, where financial management autonomy has a positive impact on the use of performance management, with no effect for personnel management autonomy (Verhoest, Roness, Verschuere, Rubecksen, & MacCarthaig, 2010).

Vining et al. (2015) observe a number of performance-related dimensions in 13 agencies in Québec during 10 years, in order to test the long-run effect of agencification on efficiency. The results of their time-series estimations indicate that agencies gain efficiency at a declining rate over time,

eventually flattening in the medium or longer term. In contrast to most studies of agencification, they resort to both objective data from annual reports (including normalized output measures) and survey data.

The work of Kim and Cho (2014) reports results on the effect of autonomy and the introduction of results control on the organizational performance of 44 executive agencies in Korea. Their data stem from in-depth interviews with civil servants in the agencies and parent ministries, as well as agencies' written rules and external reports. They find that both personnel autonomy and financial autonomy have a negative effect on performance, while contrastingly, the introduction of performance assessment and results control has a positive impact on the performance of executive agencies in Korea.

Similarly, the work of Yamamoto (2006) examines how the creation of semiautonomous agencies affected their performance in Japan. This is captured through the use of so-called retrospective surveys, which measure the intertemporal variation in performance through civil servants' perceptions. The findings point at a positive effect of operational autonomy on performance (understood as organizational effectiveness, efficiency, quality of services, and accountability), but a less clear link for other types of autonomy such as financial management, legal autonomy, or organizational structure, where institutional factors explain the different nuances.

Finally, perhaps the greatest effort in reviewing the evidence on the impact of NPM reforms in European public administrations is the abovementioned meta-analysis by Pollitt and Dan (2011, 2013) and later review by Dan (2014), which includes agencification as one of the reforms' multiple domains. The patterns they identify in a pool of around 500 works (stemming from both academics and practitioners) suggest that most works find positive effects on organizational processes and an orientation toward results and service users' needs. Yet about half of the original studies warn about unintended consequences involving fragmentation, coordination, and organizational instability. This work particularly stresses the importance of contextual aspects in defining the success of NPM reforms.

The abovementioned overview corroborates that empirical results have, for the most part shown, an inconclusive picture of the effects of agencification on the performance of public organizations, in line with previous observations (Talbot, 2004; Verhoest et al., 2004).

A number of reasons lie behind this phenomenon (e.g., see Yamamoto, 2006). Perhaps the most important one is simply the intrinsic difficulty in assessing this link at a generic level, without informing it with additional contextual variables and embedding it in its respective institutional milieu at the national level. Variables that mediate this link range from structural features of the national political system, organizational characteristics at the agency level, individual aspects related to leadership, and the qualities of those involved in the daily matters of the agency. This suggests an outstanding number of variables to be ultimately considered. A second one has to do with the divergences in the interpretation of both autonomy and performance. In the case of autonomy, it has been argued, for example, that at least six different types exist, often in tension with each other (Verhoest et al., 2004). Performance, at the same time, can include multiple aspects such as efficiency, effectiveness, quality, or accountability (Boyne 2003). Within efficiency itself, a distinction has been made between processes, outputs, and outcomes (Pollitt & Dan, 2011). An additional reason that has been put forth is the poor match between conceptualizations and measurements or classifications (Verhoest et al., 2004). To this list of obstacles, we can add others featuring less prominently in the literature. One of them is that empirical work has rarely resorted to the use of time-series evidence able to capture autonomy and performance at specific moments in time and rule out unobserved effects. Strongly related to this is the lack of studies that detect agencification itself, that is,

organizational *change*. Most articles refer to agencification, yet they only measure the number of agencies, or the mere behavior of units that were subject to agencification outside the period of interest. Finally, the literature on agencification and performance resorts overwhelmingly to the use of perception-based surveys as opposed to objective data, with all the problematic implications this has for the accuracy of measurement in the field of governance (Fazekas, Cingolani, & Tóth, 2018).

3 | IN CONTEXT: AGENCIFICATION IN GERMANY, SPAIN, AND THE UNITED KINGDOM

Processes of agencification can take on many forms, as will be illustrated by the comparative differences between our three countries of interest.

As it is commonly acknowledged in the literature, NPM-driven agencification acquired enormous momentum in the United Kingdom, aided by its government's highly centralized policy coordination capacity. At least six different semiautonomous organizational types exist in the central government's agency realm, namely: executive agencies, nondepartmental public bodies (NDPBs—subdivided into executive, advisory, and tribunal public bodies), nonministerial departments, special health authorities (subdivisions of the NHS in the UK), public corporations, and other entities (Gov.Uk, 2018). NPM-minded reforms particularly encouraged a surge in executive agencies, in contrast to the declining trend in NDPBs (James et al. 2012). Executive agencies are departmental (ministerial) bodies enjoying greater operational autonomy than line departments but which do not constitute a separate legal entity from the latter. NDPBs are separate legal entities and enjoy a higher degree of autonomy from governmental mandate than executive agencies. The United Kingdom was a frontrunner in the introduction of executive agencies. These emerged as a result of Margaret Thatcher's profound government reform of the late 1980s geared toward creating a more agile public sector. The essential structure of executive agencies has not changed since then: they are led by a Chief Executive and entrusted with reaching specific policy implementation targets set by their parent departments, against a mix of punishments and rewards determined by the performance management system defined at the central level. Although the number of executive agencies increased rapidly during the 1990s, a deagencification trend can also be observed from 1998 onward (James et al. 2012). While this trend can be mostly explained by the (comparatively) successful implementation of performance management paired with successive efforts to cut back on public spending, the creation of semiautonomous agencies continues to represent a thriving management model in the United Kingdom.

Spain's constitution envisages a centralized form of government, albeit with increasing devolved powers to both the regional and local levels. In spite of this increasing decentralization, Spain's central government retains sizable policy powers compared with other European countries. At least nine different types of agencies with varying degrees of autonomy coexist in the landscape of the Spanish public sector at the central level.¹ Among others, these include national agencies with high functional autonomy (*agencias estatales*), decentralized units (*organismos autónomos*) with generally lower levels of functional autonomy, public foundations (*fundaciones del sector público estatal*), public consortia (*consorcios del sector público estatal*), state-owned enterprises (*entidades públicas empresariales*), and other public law entities (*otras entidades de derecho público*).

Of these types, the greatest innovation introduced by NPM-reforms in Spain was the so-called “national agencies” promoted by the center-left government coalition led by José Luis Zapatero. These were introduced through the 2006 National Agencies Act, which took the form of a center-of-government single and encompassing reform, in contrast to agencification processes in other places (e.g., Germany). The 2006 Act foresaw the creation of 10 agencies and generally encouraged the

transition from other legal types into agencies in order to homogenize the existing organizational diversity (BOE, 2006). These agencies were to be managed by results under a tight contractual performance oversight by each parent ministry and all performance information collected by the newly created National Agency for Policy Evaluation (AEVAL) (Parrado, 2012). However, while a small number of the agencies survived as intended, many others were either terminated before being functional, terminated after a few years of work (such as AEVAL itself), or later retransformed into central units in the context of the downsizing policies brought by the subsequent conservative administration. In this sense, Spain displays one of the strongest policy reversals in Europe, leaving the NPM-agencification process half-baked.

Agencification in Germany happened with less intensity and momentum than in the other two countries, partly because Germany's scattered governmental structure had a large number of semiautonomous units already (Bach, 2012). In contrast to Spain and the United Kingdom, Germany holds a strongly federalized administrative organization, with large sections of policy delegated to the regional and local levels. As of 2018, a total of 18 different types of federal-level organizations exist, which can be broadly encompassed under two categories: agencies belonging to the so-called direct administration (*unmittelbare Bundesverwaltung*) and those belonging to the indirect administration (*mittelbare Bundesverwaltung*).² While many organizational types in Germany are *sui generis*, the division between direct and indirect agencies resembles a distinction commonly found in other countries, with direct agencies maintaining a degree of autonomy from their parent ministries (but not constituting a separate legal entity), and indirect agencies enjoying a higher overall level of autonomy. As explained by Bach and Jann (2010), NPM-driven agencification in Germany did not occur in the form of a centralized set of policies, nor it stemmed from a specific government coalition. It rather happened through isolated ministerial initiatives, given the high discretionary powers ministers enjoy in determining organizational structure. In this context, while a low number of semi-autonomous entities were created as a result of delegation powers coming from ministerial lines, the most frequent type of agencification happened through the passage from direct to indirect organizations. These autonomy-oriented reforms mostly responded to the need coming from specific market sectors to rely on higher credible commitment from newly emerged regulatory players, as well as greater agility.

4 | OUR MODEL: HYPOTHESES, DATA, AND RESEARCH DESIGN

4.1 | Hypotheses

Our endeavor aims to test some of the widely held beliefs around agencification in line with the academic literature on the subject. First, we use our broad empirical evidence to test the existence of performance effects of agencification, and distinguish between two different performance results: those related to administrative processes (speediness of the tendering process) and administrative outputs and outcomes (savings and tender competitiveness), probing the idea that agencification has the potential to bring improvements on all fronts (e.g., Osborne & Gaebler, 1992):

H1a *Agencification improves outcome-related efficiency in the form of value-for-money.*

H1b *Agencification improves output-related efficiency, observed by the resulting competitiveness of the process.*

H1c *Agencification improves process-related efficiency, observed by the timeliness of decision making.*

Second, in line with the interest in learning more about the timing of the potential effects of agencification and agencies' adaptation capacity (e.g., Vining et al., 2015), we test whether improvements on the efficiency of public procurement happen at all, and if so, whether these take effect in the short-to-medium term or only in the long term:

H2a *Agencification improves organizational efficiency in the form of value-for-money and this improvement takes effect in the short-to-medium term.*

H2b *Agencification improves organizational efficiency in the form of value-for-money and this improvement takes effect in the longer term.*

Finally, minding the importance of the broader context in which agencification happens and the reinforcing effects of acquiring multiple tools (e.g., McLaughlin et al., 2002), our third hypothesis is that:

H3 *The performance effect of agencification is greater in countries with higher take-up of NPM tools.*

4.2 | Causal identification strategy

Our study attempts to put these premises to test by circumventing to the extent possible some of the methodological limitations stated in the introduction. In particular, our model seeks to offer improvements by: (a) making use of objective and consistently comparative data across organizations; (b) applying the analysis to an exhaustive group of organizational transformations; (c) placing the focus on value-for-money, a rather unexplored outcome; and (d) effectively capturing the process of agencification itself, that is, the performance of agencified units against a counterfactual group, facilitating a valid identification strategy.

We capture agencification through the detection of either newly established or newly reformed public bodies that are granted a greater degree of autonomy than in their previous legal status. While this definition has the downfall of not distinguishing between the various dimensions of autonomy beyond the legal one, it has two main advantages: by relying on government's own reporting of legal autonomization, it allows an exhaustive listing of cases, avoiding any selection bias or discretion introduced by researchers; and it enables a large N strategy able to capture potential effects with greater precision. Forcing processes into a binary agencification variable, moreover, allows us to retain major changes, which is necessary in order to consider this a proper treatment. After our matching strategy with procurement data, the agencies retained largely coincide with Van Thiel's types 1, 2, and 3 (Van Thiel, 2012). To the extent possible, we exclude cases where no real

agencification took place (for example, a new agency that results from the merger of two agencies with the same former status).

As can be inferred from the previous section, the selection of Germany, Spain, and the United Kingdom responds to the fact that all three underwent some form of agencification reforms in the period of study (Hammerschmid et al., 2016), although they exhibit an important comparative difference in their levels, NPM tools take-up: high in the United Kingdom, medium in Germany, and low in Spain (Hammerschmid et al., 2016). Spain's take-up is initially high but ends up being low after a strong policy reversal.

4.3 | Data

Our analysis is based on two unique administrative datasets matched to each other. First, we collected exhaustive data on all newly created or newly autonomized agencies in Germany, Spain, and the United Kingdom between 2006 and 2016 based on official government documents, including budgets, management reports, official gazettes, national legislation, and central government websites. Together, we identified 308 semiautonomous agencies in the three countries. For the United Kingdom, we used the official website <https://www.gov.uk> as the main source. We looked for terms that identify agencies as independent, such as arm's length bodies, executive agencies, and nondepartmental and nonministerial bodies. For Spain, we also used the official government website <http://www.igae.pap.hacienda.gob.es> while also relying on the database of "Inventario de Entes del Sector Público Estatal" to identify the independent agencies. For Germany, we screened the official federal website that reports on independent agencies <http://www.service.bund.de>. The keywords we looked for were Behoerden and Bundesbehörden. Every search was accompanied with a careful qualitative filtering of false positives as explained earlier.

Second, we collected high granularity government contracting data derived from official government public procurement announcements during 2006–2016 in the three countries. Announcements appear in the Tenders Electronic Daily platform, which is the online version of the Supplement to the Official Journal of the EU, dedicated to European public procurement (DG GROWTH, 2015). The data represent a complete registry of public procurement procedures conducted under the EU Public Procurement Directives regardless of the funding source (i.e., nationally or EU-funded). All government contracts above given value thresholds³ are subject to the transparency and procedural rules of the Directives with a few exceptions (e.g., some defense contracts). All countries' public procurement legislation is within the framework of the Directives; national datasets are therefore directly comparable.⁴

This dataset contains variables appearing in (a) calls for tenders such as product specification, application deadline, or assessment criteria, and (b) contract award notices such as the name of the buying public body, the name of the winner, the awarded contract value, or the date of contract signature. For every observed tender, we have full information on contract award announcements—as its publication is mandatory—while information on calls for tenders may not be published under specific circumstances. For the three countries, we collected data on a little over 750,000 contracts awarded by central government entities, excluding all regional and local bodies for the sake of comparability.⁵

Third, agencies which were operational during the time period of our public procurement dataset had to be matched to the contracting dataset. The matching was based on agency name using a semiautomated procedure by which machine-recommended matches were manually checked in order to refine the algorithm and verify the final results. We used Levenshtein's distance partial string

matching after applying string cleaning procedures such as lowercasing and replacing abbreviations. The matched dataset contains a little over 14,000 contracts awarded by 99 autonomous agencies in the United Kingdom, Spain, and Germany.⁶

4.4 | Indicators

As anticipated, our effort explores three different dependent variables: (a) value-for-money, (b) competitiveness, and (c) decision-making timeliness. We provide a direct measure of all three of these coming from the microlevel public procurement dataset.

First, value-for-money is measured in a narrow sense, only capturing the price of purchased products. This is a suitable approach as public purchases' specifications and quality standards are defined very strictly before the tender is launched following budgetary guidelines. While we cannot fully rule out unobserved quality differences, this is minimized by the fact that we make comparisons within product categories and countries (see note in the following on our causal identification strategy). We use a direct measure of prices widely applied in the literature: percent discounts offered by the winning bidder compared with the reference price⁷ of the auction (savings henceforth) (e.g., Coviello & Gagliarducci, 2017; Fazekas & Tóth, 2018).

Second, the competitiveness of tenders is measured through the number of bids submitted to the tender. As long as collusion is barred among firms, this is a direct and reliable measure of market competitiveness (e.g., Decarolis, 2014; Fazekas, 2017). We trim the bidder number distribution at 20, as some extremely large auctions with hundreds of bidders would skew the results and typically any additional bidder beyond 20 has a little or no discernible impact on prices.

Third, we directly measure decision-making speed by counting the number of days between the tender submission deadline and the award decision (note that every time we use this metric, we control for the number of bids submitted, as more bids naturally extend the assessment process) (Fazekas, 2017).

Moreover, the subsequent analysis aims to balance the treatment and control groups (contracts awarded by autonomous agencies versus those of other central government bodies), using a number of observable contract characteristics: (a) main sector of the contract using the Common Procurement Vocabulary⁸ product classification at the highest level, leading to little over 50 product groups (e.g., architectural services); (b) contract value in EUR (using 5% of the contract value distribution as extremely high values as well as some missing records would skew the results); (c) type of procuring entity such as general government or water, energy, or telecommunications bodies; (d) procuring entity sector (COFOG classification divisions such as health care or defense); (e) year of contract award, in order to account for time-varying external shocks; and (f) the number of bidders in the model where decision-making speed is the outcome variable.

4.5 | Analytical methods

In the absence of a random assignment of contracts with respect to public body type (autonomous agency or not), we turn to the next best available analytical method: matching contracts across groups. We implement both propensity score matching⁹ (PSM) and coarsened exact matching (CEM) using Stata 14.2. We mark autonomous agencies as the treatment group and all other central government organizations as the control group (Iacus, King, & Porro, 2012; Imbens & Wooldridge, 2009). Within each country, we match on key contract and organizational characteristics as described previously: (a) main sector of the contract common procurement vocabulary (CPV) division;

(b) contract value; (c) type of procuring entity; (d) procuring entity sector; (e) year of contract award; and (f) number of bidders. Balancing the treatment and control group composition according to these covariates is expected to deliver a like-with-like comparison of contracts approximating the causal effect of agencification. While we cannot rule out the unobserved, important differences remain across the two groups, our covariates control for a large number of confounders. Also, our approach of matching central government organizations within the same time period (year) follows earlier scholarship (Vining et al., 2015) and in turn removes any potential bias coming from time-specific shocks, such as the 2008 economic crisis. Moreover, by also matching on buying organization sector and product category, we can largely avoid the pitfalls of agencies buying different things than the rest of the central government (i.e., comparing pharmaceutical purchases by the health ministry with pharmaceutical purchases by a health agency). The large scale of the dataset we have constructed allows for an efficient matching process, because for each treated contract, we have a large control group to match from (i.e., hundreds to thousands of contracts). For matching quality diagnostics, see the appendices (Tables A1–A9 in Appendix A for PSM; and Tables B4–B6 in Appendix B for CEM). We carry out both PSM and CEM but report only PSM in the main text for space considerations. CEM results are reported in Tables B1–B3 in Appendix B. The two approaches lead to the same results with only one exception, which is discussed in Section 5.

In order to test the broadest expectations regarding NPM-like agencification reforms in H1 and H2, we pool all our observations across countries, while including country-fixed effects.¹⁰ This avoids conflating the contextual realities in each case, as our theoretical reasoning suggests. Country-fixed effects account for unobserved country differences while testing, if the same hypothesized behaviors continue to operate in spite of these differences.

5 | RESULTS

Against this background, we test hypotheses H1a, H1b, and H1c by comparing the contracting performance of agencies (treatment group) to the rest of the central government (control group) according to value-for-money (savings), outputs (number of bidders), and processes (decision-making speed). Our hypotheses are partially confirmed in the three-country pooled sample (Table 1). Regarding H1a, average savings achieved significant increases ranging from 6.6% to 9.4% in the tightly matched sample, confirming our hypothesis. The effect is of very similar size and significance in CEM robustness tests in Table B1. An effect size of 2.8% points is substantial in monetary terms, as the total observed spending of agencies amounts to 78.7 billion EUR in 2006–2016. According to our calculations, Germany, Spain, and the United Kingdom saved about 1.7 billion EUR in their purchases by creating agencies.

Regarding H1b, the average number of bidders goes up from 6.38 to 6.44 in the matched samples, but this is a small and statistically insignificant effect, failing to support our hypothesis. Interestingly, the increase is somewhat larger (0.7 additional bidders) and statistically significant in the alternative, CEM, matching estimation (Table B1). However, the difference between the PSM and CEM results is largely down to different sample sizes in each of the three countries, while the country-level effects are about the same size and likewise significant (Tables 3 and B3). Overall, we find tentative but not unequivocal support for H1b.

Regarding H1c, the average decision period length (days) goes down slightly from 97 to 92 days in the matched samples with a statistically significant difference of 4.7 days, lending support to our hypothesis. In the alternative CEM estimations, the effect size becomes smaller (1.02 days) and statistically insignificant. The difference is predominantly down to different results of PSM and CEM

TABLE 1 Total effects of agencification on all three dependent variables (savings, number of bidders, and decision period length), all countries pooled, naïve and propensity score matching comparisons, including country-fixed effects

Dependent variables	Savings %	Savings %	Number of bidders	Number of bidders	Decision period (days)	Decision period (days)
Model type	Raw comparison	Matched	Raw comparison	Matched	Raw comparison	Matched
Control	7.50	6.61	7.98	6.38	119.62	97.18
Treatment	8.94	9.44	7.65	6.44	120.25	92.50
Effect size:						
Point estimate (treatment–control)	1.44	2.83	–0.33	0.07	0.63	–4.68
95% c. interval—lower bound	0.68	1.77	–0.44	–0.08	–1.59	–7.53
95% c. interval—upper bound	2.20	3.89	–0.21	0.21	2.85	–1.83
<i>N</i> control	81,593	2,784	288,915	11,331	230,486	6,996
<i>N</i> treated	2,857	2,784	11,463	11,331	8,045	6,996
Matching variables						
Contract sector (CPV division)	N	Y	N	Y	N	Y
Contract value	N	Y	N	Y	N	Y
Procuring entity type	N	Y	N	Y	N	Y
Procuring entity sector	N	Y	N	Y	N	Y
Year	N	Y	N	Y	N	Y
Nr. of bidders (trimmed)	N	N	N	N	N	Y

for Germany which we will discuss in detail in the following (Tables 3 and B3). Overall, we find partial support for H1c.

Taken together, these results tell a nuanced story. Agencies of Germany, Spain, and the United Kingdom are able to achieve considerably better value-for-money out of the same number of bids, while also marginally decreasing process/input costs. These results may also imply that unobserved tender and/or bidding quality have improved, which would be consistent with the expected higher professionalism and specialization of agencies compared with their central government peers. However, these headline effects may very well hide considerable variation across the three countries as well as agency profiles. In the following, these heterogeneous effects are explored following H2 and H3.

We test hypotheses H2a and H2b by sorting agencies according to the length of existence at the time of contracting into two groups of roughly the same size: young (less than six years) and old (six or more years of existence). We then revisit the difference in terms of value-for-money (savings %) between the control group and the two different treatment groups (Table 2). Strongly confirming H2b, and partially confirming H2a, savings improvement in old agencies is much larger than in young agencies, while both effects are statistically significant in the main models, as well as the robustness tests. For older agencies, we find that the average savings reaches 11.9% compared with

TABLE 2 Savings effects by agency age, all countries pooled, all three outcomes, naïve and propensity score matching comparisons, including country-fixed effects

Model type	Raw comparison	Matching	Raw comparison	Matching
Time horizon	Young agencies (<6 years)	Young agencies (<6 years)	Older agencies (≥6 years)	Older agencies (≥6 years)
Control	7.49	6.55	7.50	6.70
Treatment	7.96	8.51	11.50	11.86
Effect size:				
Point estimate (treatment–control)	0.47	1.96	4.00	5.16
95% c. interval—lower bound	−0.41	0.85	2.55	3.40
95% c. interval—upper bound	1.35	3.08	5.45	6.92
N control	81,593	2,784	81,593	2,784
N treatment	2095	2045	762	739
Matching variables				
Contract sector (CPV division)	N	Y	N	Y
Contract value	N	Y	N	Y
Procuring entity type	N	Y	N	Y
Procuring entity sector	N	Y	N	Y
Year	N	Y	N	Y
Number of bidders (trimmed)	N	N	N	N

the matched central government contracts (6.7% savings), a statistically significant improvement of 5.2% points (CEM finds a significant savings improvement of 6.9% points). For young agencies, the average savings increase from 6.6% to 8.5%, a statistically significant improvement of 1.96% points (CEM finds a significant savings improvement of 2.4% points). It appears that the previously observed main effect (H1a) is largely due to older, more mature agencies. Such finding strongly suggests that professionalization takes time and that the positive effects of agencification increase over time.

The second dimension of heterogeneous effects concerns the broader country context in terms of the overall take-up of NPM-type reforms (H3), expecting that the United Kingdom will conform to the traditional NPM predictions best. We test this hypothesis by revisiting all three dependent variables of H1, but dividing the treatment and control groups by country (Table 3). We find mixed support for our hypothesis. In the United Kingdom, outputs closely conform to our expectations, with the number of bidders increasing by 2.1—a sizeable and significant effect (CEM suggests an even larger improvement of 3.7 additional bidders). However, changes in savings, while positive as expected, tend to be small and insignificant: 0.28 and 1.4% points increase in PSM and CEM, respectively. The lack of significance might be due to sample size dropping to around 700 contracts only. Moreover, we find a substantial deterioration of organizational processes with average decision time

TABLE 3 Effects by three countries, all three outcomes, and propensity score matching comparisons

Dependent variables	Savings %			Savings %			Number of bidders			Decision period (days)		
	UK	DE	ES	DE	ES	UK	DE	ES	UK	DE	ES	
Country	UK	DE	ES	DE	ES	UK	DE	ES	UK	DE	ES	
Control	1.42	4.34	10.43	4.34	5.74	9.44	5.74	6.56	208.77	77.72	107.60	
Treatment	1.70	3.53	17.69	3.53	5.21	11.58	5.21	7.19	256.32	78.01	78.13	
Effect size:												
Point estimate (treatment-control)	0.28	-0.81	7.26	-0.81	2.14	2.14	-0.53	0.63	47.54	0.29	-29.46	
95% c. interval—lower bound	-2.52	-2.65	5.96	-2.65	-0.70	1.60	-0.70	0.30	27.81	-2.60	-34.49	
95% c. interval—upper bound	3.08	1.03	8.57	1.03	-0.36	2.68	-0.36	0.96	67.28	3.18	-24.44	
N control	357	1,220	1,207	1,220	7,242	1,331	7,242	2,758	567	4,360	2069	
N treatment	357	1,220	1,207	1,220	7,242	1,331	7,242	2,758	567	4,360	2069	
Matching variables												
Contract sector (CPV division)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	
Contract value	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	
Procuring entity type	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	
Procuring entity sector	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	
Year	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	
Number of bidders (trimmed)	N	N	N	N	N	N	N	N	N	N	N	

going from 208 to 256 days (and from 224 to 316 in CEM). In Germany—which embraced NPM to a much smaller degree than the United Kingdom—agencification effects are largely negligible. Both value-for-money and processes show very small and insignificant changes in all our models.¹¹ In terms of outputs (bidder number), a statistically significant change is found, although very small in magnitude: 0.5 fewer bidders (0.3 fewer bidders with CEM). In Spain—which embraced NPM strongly, but then reversed reforms in most cases—agencification has led to a clear-cut performance improvement across the board. Savings increased by 7.3% points both in the PSM and CEM estimators. Bidder numbers also increase with statistical significance, albeit only to a small degree: 0.6–1.6 additional bidders per tender in our PSM and CEM models, respectively. Furthermore, organizational processes move in the same positive direction: the length of decision making decreases by 29 days (6 days in the alternative CEM model). Given that we only compared operational agencies with their peers in the central government, extensive reversal of NPM-inspired agencification is not picked up by our estimations. In this sense, Spain provides tentative support for H3.

Taken together, it seems that an extensive and sustained move toward NPM does seem to produce positive results in line with the promises of the NPM movement, at least when it comes to outcomes and outputs. However, limited NPM reforms appear to produce mixed results, barely standing out from country-level natural variation and background noise.

5.1 | Limitations

While we have mentioned the methodological improvements that our contribution makes to the study of the impacts of agencification reform, our endeavor continues to face important challenges worth addressing by subsequent research efforts.

First, while the article's ambition is to assess whether agencification brings organizational changes that enable higher public savings related to procurement processes, it does not take a center-of-government approach assessing the overall costs and benefits of agencification. It does not consider, for example, the costs incurred in the initial setup of the agency, nor the potential added costs of establishing a result-control system, or those related to a greater fragmentation of the public sphere more generally. This is an important caveat that must be addressed in the future, even if it requires a laborious case-by-case analysis.

On the conceptual front, our study does not make justice to the actual complexity involved in agencification processes: their many types and degrees and their intricate dynamics with multiple aspects over time. The choice of a large N study limits these possibilities and reduces the conceptualization of agencification to only major legal transformations occurred at specific points in time.

Methodologically, we face the usual challenges of quasi-experimental methods. While we propose a best possible counterfactual for our identification strategy, countries vary internally in their organizational forms and how they configure their center-periphery relations, which may affect agencification in ways that are not captured here.

6 | DISCUSSION

Our endeavor enables us to engage directly with some of the key theoretical provisions laid out in the literature concerning the performance effects of agencification reforms. Consistent with the majority of previous empirical works on the subject, we are able to capture a number of specific administrative effects of agencification. These effects, however, do not speak the performance in a linear way, but rather suggest the existence of nuances and conditionalities. In particular, we do not

find consistency between the results achieved in administrative processes, outputs, and outcomes. The evidence we present suggests that while valuable outcomes, such as value-for-money, are likely to improve with agencification, there is only marginal evidence supporting the idea of process optimization. This, in turn, could help explain the persistence of mixed evidence in the literature and underline the importance of distinguishing between the multiple dimensions of performance. If traditional NPM expectations around agencification suggest an improvement on *both* outcomes and processes, the results here add to the conundrum around these beliefs, advancing the need for reformers to count on a careful balancing strategy of gains and losses of the different aspects of performance.

Our findings are particularly revealing in terms of the impact of agencification as a function of time. In line with the convincing—but scarce—prior evidence on the subject, we observe unequivocally that agencies with more years of experience in operating semiautonomously are indeed better at reaping the benefits of this (relative) independence. Due to the dichotomous nature of our variable capturing organizational maturity, we are not able to draw a full-blown time function for the effects of agencification as in other works, yet the overall higher returns for older agencies serve as an important warning for governments and practitioners pushing for quick visible results, or willing to revert reforms in accordance with (short) electoral cycles. Our findings suggest that there is a value in better understanding how time plays a role and may change the assessment of reform results.

Finally, we can cautiously endorse, through our three country cases, the generalized idea that the broader administrative context is critical in mediating the impacts of NPM-driven agencification. The biggest and clearest effects of agencification are found in Spain, the context where a wave of NPM reforms took place at once, and in the United Kingdom, the context where NPM reforms were embraced earlier, more thoroughly, and seen as most successful according to perception data. In Germany, where NPM reforms were rather marginal, the impacts were negligible. This speaks of important reinforcement effects and the pivotal role of administrative traditions as well as the consistency of administrative reform.

7 | CONCLUSIONS

While agencification has been discussed for nearly four decades, there continues to be a paucity of high-quality empirical evidence on its performance effects. In this article, we bring forth a number of findings to advance the field. We propose to circumvent a number of limitations of previous scholarship on the methodological front, yet important limitations remain and should be further tackled. In particular, fine-grained empirical strategies operating at both the agency and country levels for multiple years should be replicated and extended to more countries. While the demands around working at the very organizational level in each country, together with the challenges related to organizational heterogeneity make this a daunting task, agencification and recentralization reforms will continue to take place, increasing the worth of understanding their effects further.

ACKNOWLEDGMENT

We would like to extend our gratitude to the two thorough reviewers whose comments greatly improved the manuscript.

ENDNOTES

- ¹ For an overview of semiautonomous agency types in Spain, see Parrado (2012).
- ² See Bach (2012) and Bach & Jann (2010) for an overview of organizational types in Germany.
- ³ For example, 2016 services contracts above 209,000 EUR were published in TED. See thresholds: <https://simap.ted.europa.eu/european-public-procurement>.
- ⁴ European Commission (2014).
- ⁵ The contract data and a range of further contracting datasets are accessible at [Opentender.eu](https://opentender.eu)
- ⁶ The number of agencies matched (99) is considerably lower than the initial full sample (308) due to three reasons: (a) many agencies were no longer in operation in the 2006–2016 period for which we have procurement records; (b) there were agencies which had no discernible procurement activity either, because they do not procure independently at all or procure only small value contracts which are not included in the dataset (i.e., contract below the mandatory reporting value thresholds); and (c) we cannot rule out that some agencies were named differently in the official register of agencies and in the public procurement records (we attempted to manually identify alternative names, abbreviations, and so on for each agency to avoid missing any matches, but it is conceivable that some variants remained unaccounted for). In light of the latter point, we are confident that the final matched dataset is a nearly full population sample of agencies with at least some regulated public procurement activities, that is, all agencies other than very small ones. Agencies without independent purchasing activities cannot be considered in the analysis.
- ⁷ Reference prices are calculated in a standardized way before tenders are launched. The calculation also informs budget allocations and is designed to be independent of the bidding itself to ensure reliability of calculations. Reference prices also typically mean the maximum value the government is willing or able to pay for the products.
- ⁸ https://ec.europa.eu/growth/single-market/public-procurement/rules-implementation/common-vocabulary_en
- ⁹ We used the default profit model for estimating propensity scores, enforced common pool support and no replacement (i.e., one-to-one matching).
- ¹⁰ In practice, this means running a simple OLS regression with the agency dummy as the main independent variable of interest and including country dummies as controls.
- ¹¹ With regard to decision period length, the baseline CEM points at a significant decrease, suggesting an improvement in performance. However, upon closer inspection, this divergence between the two matching approaches is due to the considerably higher proportion of “other” and “missing” categories for the organization's main sector variable in the CEM estimates. After removing those sectors, the discrepancy disappears and the decision period impact becomes small and insignificant in CEM too (see Table B3).

Data Availability Statement

Data and supporting materials necessary to reproduce the numerical results in the article are available in Harvard Dataverse: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/X0TCZS>.

REFERENCES

- Bach, T. (2012). Germany. In *Government agencies* (pp. 18–26). London, England: Palgrave Macmillan.
- Bach, T., & Jann, W. (2010). Animals in the administrative zoo: Organizational change and agency autonomy in Germany. *International Review of Administrative Sciences*, 76(3), 443–468.
- BOE (2006) *Boletín Oficial del Estado*. Retrieved from <https://www.boe.es/buscar/act.php?id=BOE-A-2006-13011>.
- Boyne, G. A. (2003). Sources of public service improvement: A critical review and research agenda. *Journal of public administration research and theory*, 13(3), 367–394.
- Brewer, G. A. (2004). Does administrative reform improve bureaucratic performance? A cross-country empirical analysis. *Public Finance and Management*, 4(3), 399–428.

- Christensen, T., & Lægheid, P. (2007). Regulatory agencies—The challenges of balancing agency autonomy and political control. *Governance*, 20, 499–520.
- Coviello, D., & Gagliarducci, S. (2017). Tenure in office and public procurement. *American Economic Journal: Economic Policy*.
- Dan, S. (2014). The effects of agency reform in Europe: A review of the evidence. *Public Policy and Administration*, 29(3), 221–240.
- Decarolis, F. (2014). Awarding Price, contract performance, and bids screening: Evidence from procurement auctions. *American Economic Journal: Applied Economics*, 6(1), 108–132.
- DG GROWTH (2015). TED Structured Dataset (2009–2014), Tenders Electronic Daily, supplement to the Official Journal of the European Union. Brussels. Accessible at: <https://data.europa.eu/euodp/en/data/dataset/ted-csv>
- Fazekas, M. (2017). Assessing the quality of government at the regional level using public procurement data. European Commission, Directorate-General for Regional Policy: Working Papers No. WP 12/2017. Brussels, Belgium.
- Fazekas, M., Cingolani, L., & Tóth, B. (2018). Innovations in objectively measuring corruption in public procurement. In H. K. Anheier, M. Haber, & M. A. Kayser (Eds.), *Governance indicators. Approaches, progress, promise* (pp. 154–185). Oxford, England: Oxford University Press.
- Fazekas, M., & Tóth, B. (2018). The extent and cost of corruption in transport infrastructure. New evidence from Europe. *Transportation Research Part A: Policy*.
- Gov.Uk (2018) *Department, agencies and public bodies*. <https://www.gov.uk/government/organisations>.
- Hammerschmid, G., van de Walle, S., Andrews, R., & Bezes, P. (Eds.). (2016). *Public administration reforms in Europe: The view from the top*. Edward Elgar Publishing.
- Hood, C. (1991). A public management for all seasons? *Public Administration*, 69(1), 3–19.
- Hood, C., & Lodge, M. (2004). Competency, bureaucracy, and public management reform: A comparative analysis. *Governance*, 17(3), 313–333.
- Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking : Coarsened exact matching. *Political Analysis*, 20(1), 1–24.
- Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1), 5–86.
- James, O., Moseley, A., Petrovsky, N., & Boyne, G. (2012). In K. Verhoest, S. Van Thiel, G. Bouckaert, & P. Lægheid (Eds.), *Government agencies: Public sector organizations*. London, UK: Palgrave Macmillan.
- Kim, N., & Cho, W. (2014). Agencification and performance: The impact of autonomy and result-control on the performance of executive agencies in Korea. *Public Performance & Management Review*, 38(2), 214–233.
- Kydland, F. E., & Prescott, E. C. (1977). Rules rather than discretion: The inconsistency of optimal plans. *Journal of Political Economy*, 85(3), 473–491.
- Lægheid, P., & Verhoest, K. (2010). Introduction: Reforming public sector organizations. In P. Lægheid & K. Verhoest (Eds.), *Governance of public sector organizations: Autonomy, control and performance* (pp. 1–18). Basingstoke, England: Palgrave Macmillan.
- McLaughlin, K., Osborne, S. P., & Ferlie, E. (Eds.). (2002). *New public management: Current trends and future prospects*. Psychology Press.
- Moynihan, D. P. (2006). Ambiguity in policy lessons: The agencification experience. *Public Administration*, 84(4), 1029–1050.
- Osborne, D., & Gaebler, T. (1992). Reinventing government: How the entrepreneurial spirit is transforming government. *Reading Mass. Adison Wesley Public Comp*.
- Overman, S., & Van Thiel, S. (2016). Agencification and Public Sector Performance: A systematic comparison in 20 countries. *Public Management Review*, 18(4), 611–635.
- Parrado, S. (2012). Spain. In *Government agencies* (pp. 18–26). London, England: Palgrave Macmillan.
- Pollitt, C., Bathgate, K., Caulfield, J., Smullen, A., & Talbot, C. (2001). Agency fever? Analysis of an international policy fashion. *Journal of Comparative Policy Analysis*, 3(3), 271–290.
- Pollitt, C., & Dan, S. (2011). *The impacts of the New Public Management in Europe: A meta-analysis*. COCOPS Working Paper No. 3. Retrieved from <http://hdl.handle.net/1765/40668>.
- Pollitt, C., & Dan, S. (2013). Searching for impacts in performance-oriented management reform: A review of the European literature. *Public Performance & Management Review*, 37(1), 7–32.

- Pollitt, C., Talbot, C., Caulfield, J., & Smullen, A. (2004). *Agencies: How governments do things through semi-autonomous organizations*. Springer.
- Sulle, A. (2010). The application of new public management doctrines in the developing world: An exploratory study of the autonomy and control of executive agencies in Tanzania. *Public Administration and Development*, 30(5), 345–354.
- Talbot, C. (2004). The agency idea. Sometimes old, sometimes new, sometimes borrowed, sometimes untrue. In C. Pollitt & C. Talbot (Eds.), *Unbundled government*. London, England: Routledge.
- Van Thiel, S. (2012). Comparing agencies across countries. In *Government agencies* (pp. 18–26). London, England: Palgrave Macmillan.
- Verhoest, K. (2005). Effects of autonomy, performance contracting, and competition on the performance of a public agency: A case study. *Policy Studies Journal*, 33(2), 235–258.
- Verhoest, K., Peters, G. B., Bouckaert, G., & Vermeulen, B. (2004). The study of organisational autonomy: A conceptual overview. *Public Administration and Development*, 24(2), 101–118.
- Verhoest, K., Roness, P. G., Verschuere, B., Rubecksen, K., & MacCarthaig, M. (2010). *Autonomy and control of state agencies: Comparing states and agencies*. Hampshire, England: Palgrave Macmillan.
- Vining, A. R., Laurin, C., & Weimer, D. (2015). The longer-run performance effects of agencification: Theory and evidence from Québec agencies. *Journal of Public Policy*, 35(2), 193–222.
- Wynen, J., & Verhoest, K. (2016). Internal performance-based steering in public sector organizations: Examining the effect of organizational autonomy and external result control. *Public Performance & Management Review*, 39(3), 535–559.
- Yamamoto, K. (2006). Performance of semi-autonomous public bodies: Linkage between autonomy and performance in Japanese agencies. *Public Administration and Development*, 26(1), 35–44.
- Yesilkagit, A. K. (2004). Bureaucratic autonomy, organizational culture, and habituation: Politicians and independent administrative bodies in The Netherlands. *Administration & Society*, 36, 528–552.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Cingolani L, Fazekas M. The role of agencification in achieving value-for-money in public spending. *Governance*. 2020;33:545–563. <https://doi.org/10.1111/gove.12452>