The Pearson Institute Discussion Paper No. 2024-9

Purging the Police in Nascent Democracies

Monika Nalepa
Barbara Piotrowska

THE PEARSON INSTITUTE
FOR THE STUDY AND RESOLUTION OF GLOBAL CONFLICTS

Purging the Police in Nascent Democracies

Monika Nalepa*

Barbara Piotrowska†

November 5, 2024

Abstract

How do new democracies reckon with inherited authoritarian enforcement agencies? Obviously, they can disband authoritarian agencies and build new ones (thorough purge) or make former agents work for the new regime (no purge). A third way is to purge selectively: evaluate each agent, case by case. We argue, that even accounting for self-selection, such selective purges are most likely when agents are moderate in competence and ideology. When the pool of agents is less competent or more loyal to the ancien regime, selective purges give way to thorough purges; when competence and loyalty are high, "no purges" are preferred. This theory is supported with data on the operation of verification commissions in 49 sub-national regions of Poland. Building on the Most Similar Systems Design method we develop a systematic approach to case selection in instances when the number of cases is too small for OLS, yet too large to avoid cherry-picking to fit theoretical predictions.

Acknowledgments: The authors are grateful for comments, criticisms and suggestions to Jack Paine, Joseph Rugiero, Olga Gasparyan, Georg Vanberg, Joe Wright, and Yusuf Magia. All problems are the authors' responsibility. Ryan Gibbons provided superb research assistance.

*The University of Chicago, Department of Political Science, mnalepa@uchicago.edu

†King's College London, Barbara.Piotrowska@kcl.ac.uk

1 Introduction: Purge or reform?

The protection of the rights of life, liberty, property, and contract is a fundamental function of the state, even according to the most minimalist conception (Nozick 1978; Mack 2018). Governments must ensure that the agencies offering protection, such as the police, security services (Hassan 2020), and courts function effectively. But how can new governments recovering from periods of authoritarian rule reform, re-purpose, or replace their security agencies without temporarily surrendering these functions of the state? This paper reconstructs this dilemma facing policymakers in the aftermath of the transition to democracy.

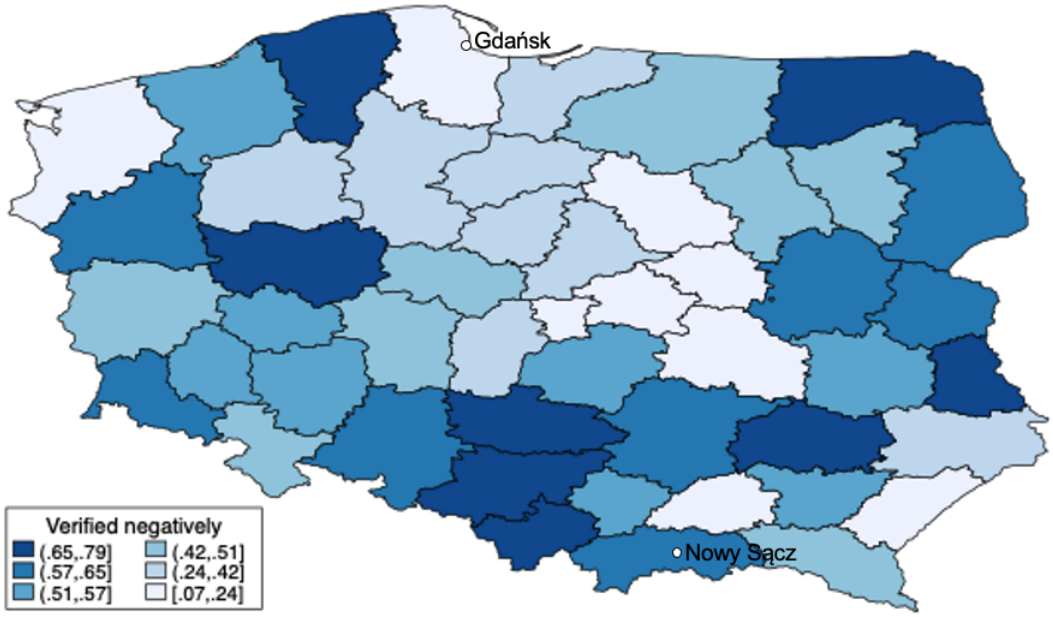
Newly democratized states confront a gamut of choices of how to deal with the personnel of compromised security agencies. The seemingly obvious answer to the question of who should be fired from authoritarian enforcement agencies— "the bad apples"—is ambiguous. It could refer to firing incompetent officers. Alternatively, it could refer to removing those appointed to their positions for their ideological loyalty to the authoritarian leaders. The former are undesirable because they are incompetent and retaining them offers no value to the new state. The latter are undesirable because they focused on prioritizing the interests of the authoritarian state rather than serving the interests of its citizens. Among the former staff of the enforcement agencies are also those whose competence can become useful to the new democratic regime and who were not particularly loyal to the autocrat. Who specifically these desirable agents are is not obvious from the outset and requires scrupulous vetting. Hence, in deciding how to deal with the former enforcement agencies, the new democracy has a third option. Beyond (1) purging everyone in a "thorough purge" or (2) doing nothing ("no purge") they can (3) try to carefully vet the personnel and "pick out the bad apples." Under what conditions will they conduct such a selective purge? And under what circumstances will careful vetting give way to a thorough purge rather than to doing nothing?

To answer these questions, we use a simple formal model that reconstructs the decision-making process of a body tasked with vetting former security officers for reemployment eligibility. Beyond clarifying how purge decisions are made, our theory makes three additional

contributions. First, it stipulates when one has to account for the fact that former security agents may select out of screening. To describe when such selection effects arise we extend the simple model into an incomplete information one. Second, this extension allows us to dispel with the conventional wisdom about how vetting commissions operate. Our third contribution is to empirical methods. We adapt Most Similar Systems Design, an approach used in case study research, to hypothesis testing in circumstances where the number of cases is too small for OLS, yet too large to escape an accusation of cherry picking cases for comparison.

To corroborate our theory, we use granular data from post-communist Poland, a former socialist state that, at the critical juncture of communist collapse, had to purge its security and enforcement agencies. Poland is an ideal setting for testing our theory of how policy-makers resolve this dilemma, as it delegated each verification decision concerning each employee of the former secret police to one of 49 Regional Verification Commissions (*Wojewodzkie Komisje Kwalifikacyjne*, RVC), permitting us to study 49 purge decisions, as exhibited in Figure 1.

Figure 1: Geographic distribution of the proportion of secret service rejected in the screening process



We highlight two features (one theoretical and one empirical) of the map that link to our contributions in this paper. On the theoretical side, purges of secret police officers do not conform to expectations from the transitional justice literature.[1] A quick scan of the the map indicates that areas with the highest authoritarian repression did not experience harshest purges. The commission in Gdansk (in the north), a site of significant communist-era repression (Ekiert 1996), was among the most lenient, allowing more than 90% of former secret police officers to keep their positions. On the other end of the spectrum, areas with little to no authoritarian repression used purges less sparingly. For example, Nowy Sacz (in the south), with hardly any dissident activity (Interview 2023), had one of the harshest vetting commissions, banning almost 70% of former secret police officers from working in the enforcement agencies. A telling feature of the Nowy Sacz commission was that, even though it had the capacity to do so, it refused to interview, for vetting purposes, a single officer, relying on personnel files and personal knowledge of the officers alone.

In short, secret police officers who resorted to extreme violence and persecuted the anti-authoritarian opposition may have "deserved" to be punished by members of the incoming democratic regime, but this does not seem to be the guiding principle of verification commissions. According to some of the democratic reformers, retribution was not their focus: "While implementing this 'revolution', it was key to abandon any desire for revenge, because such revenge instead of hurting its intended target would end up consuming us [the former anti-communist opposition]" (Wszolek 2019, p. 186). If past repression against political dissidents does not explain vetting decisions across regions in post-communist Poland, what does?

We argue that while repression is not irrelevant as a purge criterion, in contrast to transitional justice reasoning, it is used as a proxy of suitability for the democratic police force rather than an indicator of an officer "deserving" to keep his or her post. A history of repressing dissidents is not immediately damning to a police department but can be

[1] See (David 2003; Horne 2017)

mitigated by the competence of agents in that department. Purge decisions are far from being uni-dimensional: in addition to the extremism of the officers, they take into account officers' competence, best understood as skills useful for policing in a state that resepcts the rule of law.

Second, on the empirical side, as shown in Figure 1, vetting commissions exhibited considerable variation in their decisions. This variation in verification outcomes suggests opportunities for testing our theory with statistical tools. However, 49 regions (called *wojewodztwa*) is still too small a number for a reliable large-N analysis. We take this challenge as an opportunity to develop a new approach to Most Similar Systems Design (MSSD). MSSD along with paired comparisons has traditionally been used for theory generation (Gerring 2016). Sydney Tarrow compared the method to experimental designs, as it allows the researcher to isolate the impact of a "single variable or mechanism on outcomes of interest" (Tarrow 2010, p. 224). To our knowledge, the method has never been used for testing formal models. Beyond applying MSSD, we develop a method for case selection that ensures that the cases considered are indeed *most* similar by leveraging rich quantitative data to calculate the Euclidean distance between the observations.

This paper is organized as follows: The next section positions our theory in relation to two bodies of literature: first, research on the loyalty competency trade-off, and second, the scholarship on purges. Both of these literatures center on authoritarian regimes. The following section presents a decision theoretic model that adapts the loyalty-competency trade-off framework to the context of regime transitions. We show that recasting this familiar trade-off in this new setting allows us to productively analyze the dilemma facing reformers of enforcement agencies (Egorov and Sonin 2011). The following section tests the theoretical robustness of this model by examining contexts under which we should account for selecting into verification by the officers themselves. The remainder of the paper is devoted to the empirical corroboration of our findings. In addition to analysis of newly assembled Institute of National Remembrance (IPN) data, it develops the a method of using MSSD for selecting

cases.

2 Authoritarian purges versus post-authoritarian verification

The past decade has seen renewed interest in how autocrats use their enforcement agencies to maintain power (Blaydes 2018; Hassan 2017; Hassan, Mattingly, and Nugent 2022; Greitens 2016; Scharpf and Glassel 2020). In 2016, Sheena Chestnut Greitens (Greitens 2016) argued that autocrats' organization of their repressive agencies is a response to the kind of threat they face. If they are primarily concerned with revolution from below, they will centralize secret police forces. Conversely, if their main concern is a coup d'etát, they will fragment their security forces and sacrifice efficiency in intelligence gathering to prevent a lateral challenge to their rule. The ensuing structure of enforcement agencies will have implications for the competence of officers employed in the secret versus uniformed police. In Argentina, where transitions in power between one junta to the next occurred through coups, the security agencies attracted incompetent agents (Scharpf and Glassel 2020). According to their argument, in Argentina, the career incentives in alternative hierarchical organizations, such as the military, made it challenging for mediocre agents to advance there. The ambitious, yet talent-lacking, officers could further their careers by undertaking arduous work required by secret police.

In states populating the Soviet bloc, however, the predominant threat to communist rule came from below. Accordingly, the secret police was highly centralized (Thomson 2024). The Polish Communist secret police agencies were also surprisingly competent relative to the uniformed Citizens' Militia (the police force). According to Dudek and Paczkowski (2005), the quality of Polish security service officers improved significantly in the 1980s compared to the early years of the regime. In Appendix B.1, we provide evidence for Poland specifically indicating that secret police employment was more attractive than the alternative career

paths. Over time, and especially by the early eighties, secret police officers became more competent.

Following the transition from authoritarian rule, the goal of new democrats is to surround themselves with competent and loyal agents. However, attaining both of these attributes simultaneously in not always possible. An advanced political economy literature highlights the trade-off between loyalty and competence in authoritarian regimes and posits that authoritarian stability lies in its successful resolution (Egorov and Sonin 2011; Zakharov 2016).

According to Egorov and Sonin (2011), dictators, have to balance their demand for loyal agents with the need for skilled ones. In the words of these authors,"the very competence of the vizier makes him more prone to treason" (p. 904). They explain how both competent, albeit less loyal agents, and loyal but less competent agents can exist on the autocrat's payroll.

This idea is illustrated in Figure 2, panels (a) and (b). The ideal points of actors–a Dictator, an Incompetent and Loyal agent and a Competent but Disloyal agent [2]–are represented in a hypothetical issue space. Following intuitions from delegation models (Gehlbach 2021), we assume that the need for agents stems from the fact that they implement policy. The way that competent and incompetent agents differ is that the former can effectively detect signals about the direction of random distortions to policy. In panel (a), this is represented with ϵ. Any agent assigned a policy to implement will strive to implement it at the location of their ideal point. A competent agent will implement it exactly at his ideal point because even if such implementation is perturbed by a shock (ϵ), he is able to observe it and correct for it. Incompetent agents cannot observe the direction of the policy perturbation and so even when they attempt to implement policy at their ideal point, their attempt fails. In light of this, the dictator is better off employing competent agents even when the ideal point of such an agent is further away from the Dictator than that of the loyal agent. Just as in Egorov and Sonin's (2011) model, a "competent vizier", although he is not as loyal to sultan,

[2] We assume, for now, that the supply of both Loyal and Competent agents is negligible

may be retained precisely because of his competence.

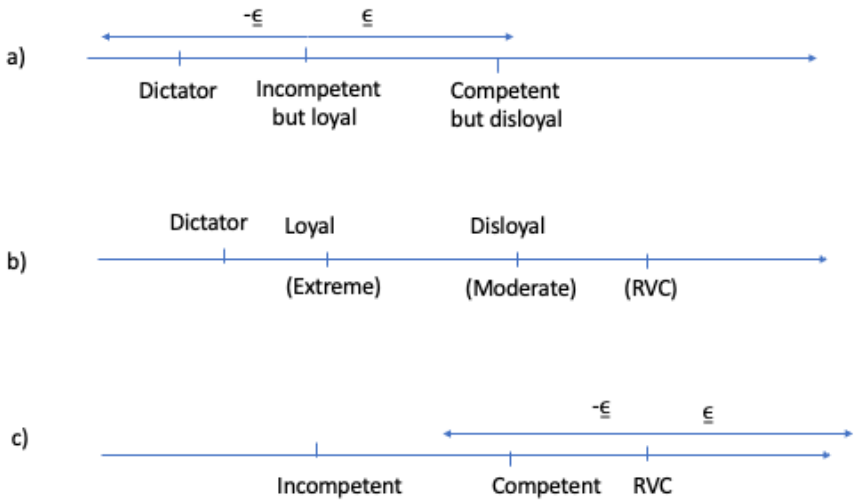


Figure 2: Reinterpretation of the loyalty-competency trade-off in post-authoritarian setting

Next, to understand the loyalty-efficiency trade-off in the context of a regime transition, consider panel (b) of Figure 2. This panel features an additional actor, the RVC, with an ideal point intuitively positioned at the opposite end of the issue space from the Dictator. To represent the ideal points after the transition, we have put parentheses around the relevant actors after the transition. This way it is easy to see that an agent disloyal agent to the dictator is actually "Moderate" from the point of the RVC while an agent who was loyal to the dictator would be considered "Extreme" *from the perspective of the RVC*.

Finally, panel (c) of figure 2 illustrates how the context of post-authoritarian purges may allow democratic authorities to trade-off between loyalty and competence altogether: agents with expertise and weaker loyalty to the outgoing authoritarian regime are both more loyal to the new democratic regime than the incompetent agents *and* more skilled at implementing policy than newly recruited agents appointed by the RVC would be. This type of agent is labeled as "Competent" in panel (c). The problem of post-authoritarian purges is not one of

avoiding the trade-off between loyalty and competence, as in authoritarian regimes, but one of screening or selecting the moderate and competent types from a pool that also includes extreme and incompetents types.

Two points are worth noting here. The first, also represented in panel (c), is that if the RVC were to purge all agents, it could hire as their replacements untrained agents with preferences identical to the RVC but with no training. This lack of training of newly hired agents is also represented as the perturbation ϵ, which can extend in either direction of the RVC. Provided ϵ is sufficiently high, as it is in Figure 2 the RVC would prefer to retain the agent of the previous regime to hiring a brand new force.

The second point extends back to the authoritarian period. Presumably, if the supply of competent agents is sufficiently abundant, autocrats would have had the opportunity to hire agents who are both competent and loyal. After the regime transition, this would affect vetting decisions by expanding the pool of agents that the RVC must vet to include not only competent and moderate agents but also incompetent and extreme agents, as well as competent and extreme agents. Moreover, to the extent that competences expected of police agents before and after the transition vary, the pool will also have incompetent moderate agents.[3]

In sum, the issue of loyalty and competence evolves during the transition from being a trade-off necessary for maintaining authoritarian stability to becoming a selection problem characteristic of purge decisions.

This brings us to the next body of literature that relates to our contribution - the literature on authoritarian purges. Starting with Jun Sudduth (2017), who theorizes and empirically demonstrates that purges occur when rival elites are temporarily weak rather than strong, this body of work challenges conventional views in the literature on authoritarianism. Recall that the line of work pioneered by Sheena Chestnut Greitens theorized autocrats as structuring their security agencies to stave off threats to their power. Sudduth

[3] Such agents would have been appointed despite disloyalty, for competence in an authoritarian regime; this competence, however, would not carry over to the democratic regime.

shows that purges cannot be interpreted as a straightforward extension of this strategy. A purge aimed at rival elites when they are strong could provoke them to entertain a coup in response to a purge. Thus, even though a purge against strong elites is most useful, it is also less feasible than a purge against weak elites who hope to get stronger one day. The latter purge proves to be more effective in helping dictators consolidate their power as long as it is not so extensive to destroy the elite's hopes of recovering their strength one day. In her analysis, Sudduth characterizes a distinctive feature of purges and demotions relative to show trials of disappearances: Whereas the latter are final, the former, as also work by Mattingly (2024) demonstrates, can be reversed, suggesting that it is well within the dictator's powers to fine tune the severity of purges. This feature of Sudduth's model makes it a poor candidate to use as a template for post-authoritarian purges, however, because democracies being entirely non-violent regimes cannot calibrate the severity of purges in the same way. Moreover, Sudduth in her theory assumes that the elites who are threatened by being purged, not only know if they will rebound in period 2 relative to period 1 ($p2 > p1$) but also by how much they will strengthen their power (with or without purges). While we are not disputing that this is a reasonable assumption in relatively institutionalized authoritarian regimes, it is definitely too strong for the transient context in which post-authoritarian purges are carried out.

Another formal model of authoritarian purges by Montagnes and Wolton (2019) uses a principal-agent framework (one principal to many agents) to examine the effect of purge decisions made by autocrats on agents' efforts, where agents could be any members of the regime—from enforcement to bureaucrats—as long as they are not elites as in Svolik (2012)'s model. The parameters of the purge are set by choosing a level of violence (purges are violent affairs, but not uniformly so) and two purge thresholds—one for agents who failed at some task, and one for agents who succeeded. In one of the key results of the paper, congruent agents exert effort in response to the autocrat setting the threshold of purging in response to failure sufficiently high. Here, congruent agents always exert effort, suggesting that purges as

a strategy of accountability work. Even though, ostensibly, this is a model of authoritarian purges, there are many elements of the Montagnes and Wolton (2019) model that we can use in our theory, such as the idea of congruence with the principal who decides purge levels and what makes purges in some circumstance costlier than others. However, in contrast to their model, our principal does not have an opportunity to observe the effort exerted by an agent, even if indirectly. Also, due to the democratic constraints of our problem, selecting the level of violence is not an option here.

In a most recent contribution, Ketchley and Wenig (2023) discuss purges in Egypt's post-colonial period, which accompanied the 1952 junior-officer-led coup. Focusing on military turnover, the authors document how, in the coup led mostly by junior officers, a pattern emerged in how purges were conducted: Instead of a thorough purge that may have been expected from the violent and revolutionary nature of the coup, according to the authors' analysis of over 600 biographical registers, Gamal Abdel Nasser, the leader of the young plotters, made purge decisions "consistently with a threat-competence calculus." Certain elites, notwithstanding their degree of entrenchment in colonial rule, were allowed to maintain their positions because of the coup plotters' belief in their high levels of competence. The reasoning behind decisions to fire or rehire personnel from the colonial military closely aligns with our theory of post-authoritarian purges.

In sum, although we are able to draw on certain elements of the formal and quantitative literature on authoritarian purges (see also Bokobza et al. (2022); Crabtree, Kern, and Siegel (2020); Goldring and Matthews (2023)), our choice to focus on decisions pertaining to low-level enforcement agents rather than the leaders and post-authoritarian time rather than authoritarian period compels us to develop a new theory.

3 Assumptions and Theory

To model the decision faced by new democratic politicians of how to deal with members of an authoritarian security agencies, we build on the workhorse model of American politics scholars studying bureaucracies: delegation theory (Epstein and O'Halloran 1999; Huber and Shipan 2002; Callander et al. 2008). As a point of departure from this literature, the new democrat's decision centers on whom to purge from the authoritarian state agencies rather than whom to appoint. Hence, the problem of whether or not to fire a law enforcement agent is the reverse of the delegation problem: instead of choosing to delegate, the politician in a new democracy must decide whether to retain an existing agent previously appointed by a different - and authoritarian - principal.

3.1 The decision-theoretic model

We can model the new democrat's dilemma as a decision-theoretic problem, where a government official (such as an RVC chairman) makes a choice about a former law enforcement agent. The agent can be competent or incompetent; we assume he is competent with probability $c \in (0, 1)$. Independent of his level of competence, the agent can also be moderate or extreme; he is moderate with probability $m \in (0, 1)$[4].

Using the insights of Nalepa (2022) and Gehlbach (2021) and to simplify analysis, we assume that if the RVC could perfectly and at no cost observe the moderation and competence of the agents of the enforcement agencies, it would retain those who are simultaneously competent and moderate.

Although the RVC knows only the aggregate distribution of competence and loyalty in its

[4] The difference between competence/incompetence and extremeness/moderation is similar to the difference between skills versus attitudes. For instance, work in industrial espionage or bureaucratic corruption demanded skills that could be repurposed to fight white-collar crime. These will be exactly the kinds of agents that politicians responsible for the reform of the police agencies will want to retain. Ideology includes an understanding and respect for human rights. E.g., the Polish police in the 1990s were asked to attend workshops on the European Convention on Human Rights because ideological extremeness was seen as a clear impediment to serving the new democratic state once the nature of security service work has shifted from serving the state to protecting the citizens (Piotrowska, Szkurłat, and Szydłowska 2024).

district, it can pay a cost, s (representing the cost of a selective purge) to learn the loyalty and competence of the individual agents. This cost is interpreted as taking the time to examine the personnel files of secret police officers. These files contain relevant information about their performance, whether they received awards or sanctions, their tenure on the job, as well as career trajectory, transfer history, and education acquired before and during their service in the secret police.

In the first period, Nature determines if the agent is competent and whether his ideal point is moderate or extreme.

In the second period, the RVC chooses one of three actions $a_R \in \{f, v, h\}$:

1. Thorough purge (f), where the RVC fires all agents and trains replacements at a cost of $t \leq 1$[5]

2. Selective purge (v), where the RVC pays cost s to learn the competence and ideal point of the agent, and only keep the moderate and competent agent, while firing all other types.

3. No purge (h), where no agents are fired and no new information is learned.

As explained above, the goal of the government officials from the RVC is to employ an agent who is competent and moderate. The value of such an agent is 1. Otherwise, the government official's payoff is 0. Hence, the payoff to a thorough purge (1) is $1 - t$, where t is the cost of recruiting and training a new agent. Here, the RVC fires everyone and replaces them with agents who are initially also incompetent but whose preferences are aligned with the RVC. Similarly, the payoff to a selective purge (2) is $cm + (1 - cm)(1 - t) - s$, which reduces to $1 - (1 - cm)t - s$. With no purge at all, given the probability of a competent agent is c and the probability of a moderate agent is m, the payoff from "no purge" (3) is cm. Recall that, without careful screening, the officials do not know the realization of c and

[5] In terms of Figure 2, we assume the replacements shares preferences with the RVC but lack competence. Consequently, this agent cannot observe and correct for policy shocks ϵ until they have received training

m of each agent, but rather the distribution of c and m in their district and base their purge decision on that distribution.

Given the parameters of the model above, we can derive the conditions under which a selective purge is the optimal choice.

Proposition 1. *A selective purge is preferred to a thorough purge, which is in turn preferred to no purge when*

$$\begin{cases} cm < 1 - t \\ s < cm * t \end{cases} \tag{1}$$

Proposition 2. *A selective purge is preferred to "no purge", which is in turn preferred to a thorough purge when*

$$\begin{cases} cm > 1 - t \\ s < (1 - t)(1 - cm) \end{cases} \tag{2}$$

These conditions are nonlinear in cm and the condition of each of the propositions determines whether the "runner-up" choice is "no purge" or a thorough purge:

1. $cm < 1 - t$ (thorough purge preferred to "no purge")

2. $cm > 1 - t$ ("no purge" preferred to thorough purge)

For any given level of training cost, there is a cutpoint in terms of cm (the joint probability of a competent and moderate agents) that determines which of the constraints on s (the cost of carefully reading personnel files) is binding. This cutpoint is also the inflection point. In Figure 3 this inflection point is set at $t = \frac{1}{2}$ Prior to reaching that point, the range of s for which a selective purge is optimal is increasing in cm, beyond that point, it is decreasing. A selective purge is easiest to carry out right at the inflection point, but its optimality decreases as one moves away from it. However, to the left of the inflection, a thorough purge becomes more attractive than a selective purge, while to the right of the inflection point "no purge" becomes more attractive to a selective purge.

3.2 Discussion

The above findings are intuitive. Note that for cm (represented on the horizontal axis in Figure 3) to be high, both competence and moderation must be high. The fact that for such high values, the price for a selective purge to be optimal must be very low makes sense: why pay the cost of a selective purge when almost all the agents of enforcement are desirable? Unless the costs of training new staff are very low, it makes sense to just preserve all agents in their positions, implementing "no purge". However, a decrease of cm (which happens when either c or m falls) erases the advantage of "no purge". For a high cost of screening s, moving to the left on the horizontal axis, i.e. decreasing the expected competence and moderation of the agents, makes a thorough purge more attractive than retaining all former agents. Similarly, moving from top to bottom along the vertical axis (i.e. decreasing the cost of screening), the attractiveness of a selective purge increases relative to a thorough purge for low values of cm and relative to "no purge" for high values of cm. In a nutshell, the predictions of the model diverge depending on whether the proportion of competent and moderate agents is high or low. When the proportion of both or either is low, the alternative to a selective purge is a thorough purge, but when it is high, the contender is "no purge".

4 Robustness check: a signaling model

The advantage of the model presented in section 3.1 is its simplicity and intuitive solution. As with all simple models, however, this one too misses some of the complexity of the real world, probably the most troubling of which is illustrated below.

The map in figure 4 displays the proportion of officers who applied for verification relative to the total number of officers hired in 1985[6]. It suggests that perhaps a spot in the secret police was not uniformly lucrative across all wojewodztwa.

Suppose then that, contrary to our decision model, officers had an opportunity to *select*

[6] We use 1985 as a benchmark because it provides sufficient time before the transition, ensuring that officers had not begun leaving strategically in anticipation of the regime change.

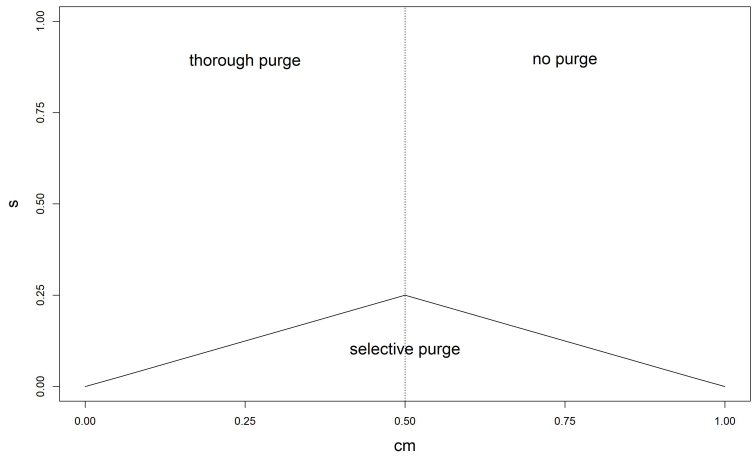


Figure 3: Thorough, Selective and "no purges" as a function of competence and loyalty (horizontal axis), and cost of selective purge (vertical axis) for three different costs of training (dashed vertical lines) of new agents

into verification. In other words, the decision to apply to be considered for verification was not random, but itself a function of additional factors, critical among them opportunity costs of the potential applicants. Perhaps, by failing to account for differential opportunity costs, our model ignores these selection effects and leads us to unreasonable expectations? In this section, as a robustness check of sorts, we generalize the decision model above to incorporate decisions of former secret police officers whether to apply for verification.

Introducing this additional action of the officers not only makes the model more realistic[7], but it allows us to evaluate conventional wisdom about the way purge commissions "discourage" undesirable officers from applying for reemployment.

A natural way to model selection into vetting is by using a signaling model, the details of which can be found in Appendix A.1. In this extension, each officer has private information about their own qualifications and political background and in light of this information must decide whether to apply for vetting by the RVC. If they do not apply, they may seek out employment opportunities in the growing private sector, such as in private security or in new firms, where we assume political loyalties to the ancien regime do not pose a problem.

[7] which by itself should be a reason for enriching a model (Clarke and Primo 2012)

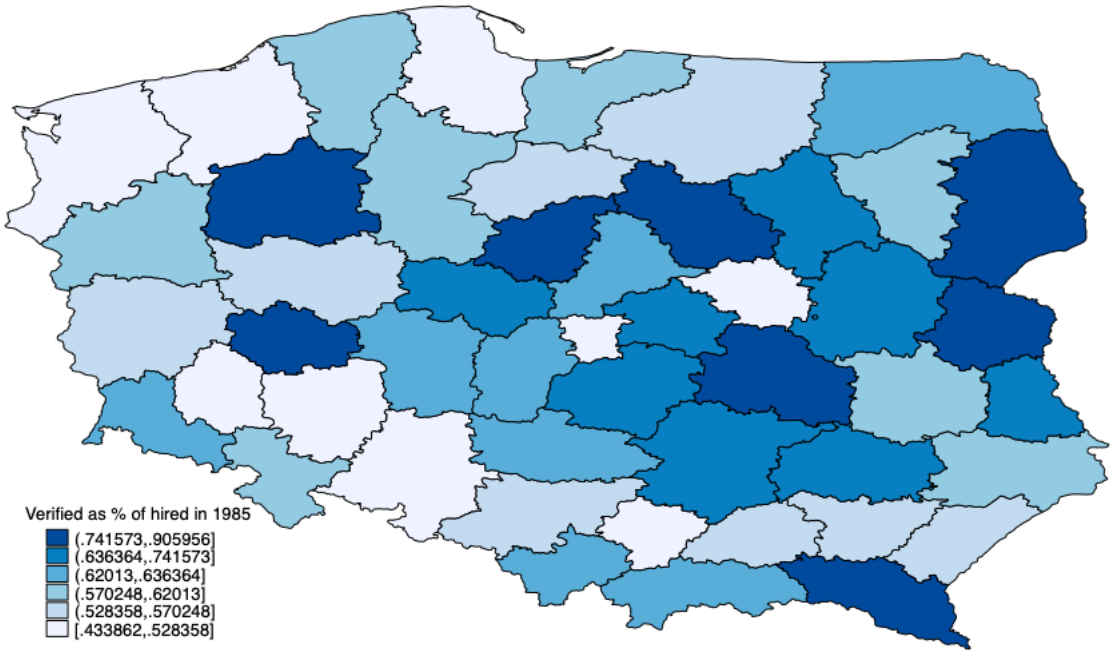


Figure 4: Proportion of officers who applied for verification (1985 benchmark)

If they apply for positions in the democratic security forces, they are vetted by the RVC.

The RVC has limited information, knowing only general trends in competence and loyalty but not the specific details of each applicant. To get more information, the RVC can pay to examine the personnel files of applicants, which may or may not contain evidence of incompetence or extremism.[8] If the RVC chooses to fire an undesirable officer, that officer's future job prospects are worsened due to time wasted during the vetting process. We assume, however, that the desirable type of officer's job prospects are relatively no affected by awaiting results of the verification process. Thus officers' decision to apply for verification serves as a signal to the RVC and is risky for undesirable officers, as it may cost them job opportunities. The RVC, in turn, must decide whether to hire without vetting, invest in costly vetting, or fire the applicant. The fired applicants need to be replaced with new recruits whose training is costly.

The game's equilibria are characterize fully in Appendix **??**. Here, it suffices to say they

[8] Importantly, we assume some files may have been destroyed during the regime transition, meaning the vetting process is not guaranteed to reveal undesirable traits.

fall into has three potentially possible classes: pure separating equilibria (in which only one type of officer applies for verification), semi-separating equilibria, where both types apply but one type does so with a higher probability, and pooling equilibria, in which all the types apply at the same rate.

We start the analysis with the pure separating equilibria, which, incidentally, reflect the conventional wisdom about how verification commissions operate. According to this argument the mere act of setting up such commissions will dissuade undesirable types from applying because they know that evidence against them is in the hands of the commissioners. When the RVC conducts a selective purge, knowing there is no evidence against her, the desirable type agent is not constrained in applying, but undesirable type refrains from applying. Can such an equilibrium exist?

It can not. To see why, suppose to the contrary that such an equilibrium exists. If that were the case, the RVC upon seeing an application, would know with certainty the applicant is desirable and the RVC's best response would be to hire the applicant without detailed screening, which incurs the costs of vetting. However, the undesirable agents' best response to such a strategy of the RVC would be to start applying in droves as there is no risk they would be uncovered. And so the separation and the equilibrium falls apart. Consequently, no separating equilibrium exists where only desirable types apply.

It is even easier to show that the reverse scenario—where only undesirable agents apply and desirable ones do not—does not hold. If that were the case, the RVC would have to assume that any application is coming from undesirable agents. However, the payoff for hiring undesirable agents is worse than the payoff from firing them, so the RVC should fire them. For the officers, in turn, the payoff from applying and being fired is lower than the payoff from not applying all (recall that for this type, the value of the outside option diminishes over time). Hence, no pure separating equilibrium exists.

Having ruled out both types of pure separating equilibria, we are left with two other classes: semi-separating and pooling equilibria.

In semi-separating equilibria, the commission partially learns the applicant's type. In pooling equilibria, no distinction between types is made as all types of officers apply.

The existence of pooling equilibria, where all agents apply at the same rate, regardless of their type, is critical for us for two reasons. The first is theoretical: if most of the parameter space of a game is filled with outcomes associated with pooling equilibria, this can be taken as lack of evidence of self-selection into screening. The absence of such self-selection is also an implicit assumption of our simple decision-theoretic model. Even if pooling equilibria fail to fill the entire parameter space but cover the part that corresponds to our scope conditions on the ground, we can be confident in using the decision model to guide our expectations regarding Poland. The second reason pooling equilibria are important is empirical: outcomes associated with pooling equilibria can be tested with aggregate data. Results from solving our signaling model are summarized in Figures 6a and 6b. Jointly, they indicate that pooling equilibria occur under a wide range of conditions, particularly when the cost of training of new loyal agents is high. This is because, under those circumstance, undesirable agents are more likely to apply for verification. In this case, the RVC is incentivized to keep even undesirable agents, leading to a situation where both desirable and undesirable agents apply.

Hence, the signaling model offers two important insights. First, it challenges the common belief that vetting commissions can easily separate undesirable candidates from desirable ones. Second, it demonstrates that, for sufficiently high cost of training of new loyal agents, pooling equilibria cover a wide parameter space.

The next figure we present shows clearly that the simpler decision-theoretic model captures all the key dynamics of the purges.

We present the decision of the RVC as a function of cm, the proportion of desirable agents (on the horizontal axis) and s the costs of vetting.

This figure is strikingly similar to Figure 3. Just as in the decision model, there is also a triangular area (in red, demarcated by two lines—one corresponding to the constraint defining what makes vetting better than hiring and one corresponding to the constraint

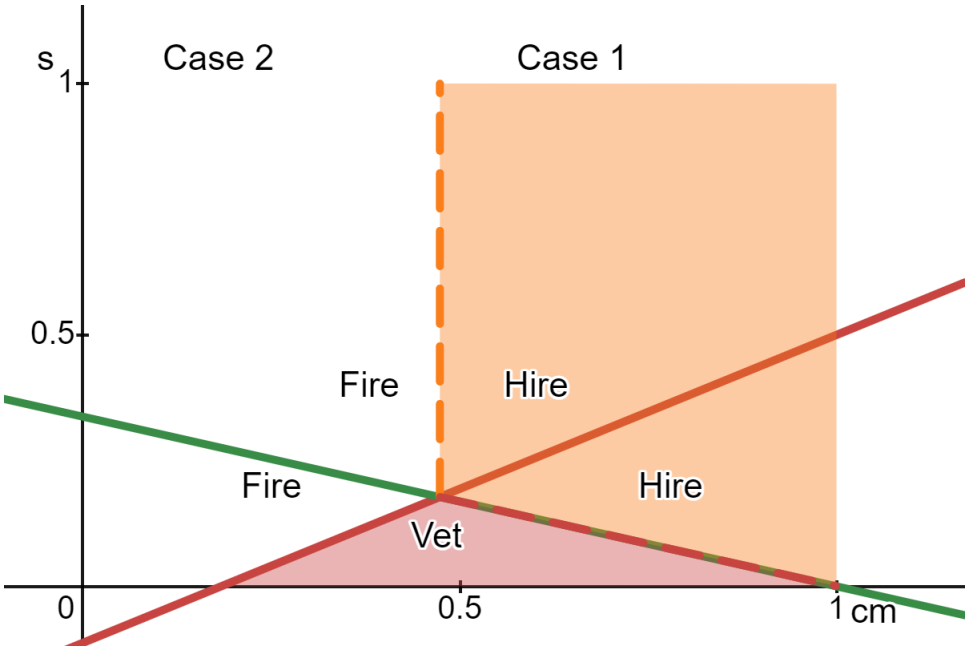


Figure 5: When is vetting, firing, or hiring an optimal action of the RVC, assuming $t = .5, d = .95, q = .75$ (see Appendix **??** for definition of parameters)

defining what makes vetting better than firing). Also, just as in the decision model, it is easiest to satisfy the joint conditions for middling proportions of desirable agents (where the red triangle reaches its peak). Above the red triangle, the optimal choice of the RVC is either "hire" (in the orange area) or "fire" (everywhere else). What determines the preferability of hiring versus firing is whether the proportion of desirable agents is smaller or greater than the inflection point, which is a function of t.[9] This is another similarity to the decision model: as the costs of training associated with firing increase, the peak of the triangle shifts to the left and the hiring area expands at the expense of the firing area. Figure 1 in the Overflow Appendix corresponds to this situation. A link to the Overflow Appendix is at the bottom of Appendix C.

The upshot of this, is that in the terms of the best response of the RVC, the pooling equilibria match up very intuitively with the decision model. If the parameter space associated then with the semi-separating equilibrium is relatively small compared to the pooling

[9] Concretely, defined in terms of the signaling model, the value of the inflection point is $cm = \frac{d-t}{d}$ (See Appendix A.4.

equilibrium, our robustness check can be interpreted as affirming that the decision model is sufficient.

The formal proposition describing when pooling equilibria in which both types apply while the RVC takes one of the action fire, hire, or vet is in Appendix **??**, following the solution for both to pooling and semi-separating equilibria. Intuitively, semi-separating equilibria prevail when the opportunity costs associated with applying for verification by undesirable types are sufficiently high, because when the opportunity cost is low, incompetent and extreme types will be more likely to refrain from applying for verification. The question then becomes when is the opportunity cost high enough to create such semi-separating effects and when is it so low that pooling prevails. The answer to this question is captured in the final figure of our theoretical section, below.

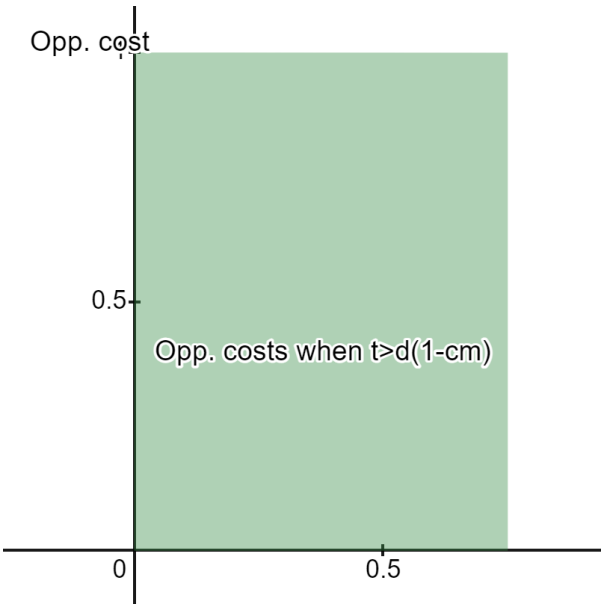
The shaded areas in Figures 6a and 6b capture the area that the opportunity costs must fall into for the pooling equilibrium to occur.[10] Each subfigure corresponds to one of the two subcases considered in the appendix where equilibria were found (low training costs on the right and high training costs on the left.[11]

We see that for high training costs (the left panel), there exists a pooling equilibrium for the entire parameter space of the opportunity costs. This can be interpreted in an intuitive way: lucrative employment opportunities for the undesirable types are more likely when it costs more to train their replacements. In line with our reasoning above, the decision model from section 3.1 is sufficient for explaining purge decisions here because the predictions for optimal actions of the RVC stemming from the pooling equilibria are exactly the same as in the case of the decision model. For low training costs, however, the shaded area, corresponding to pooling equilibria, only covers select (and low) values of the opportunity

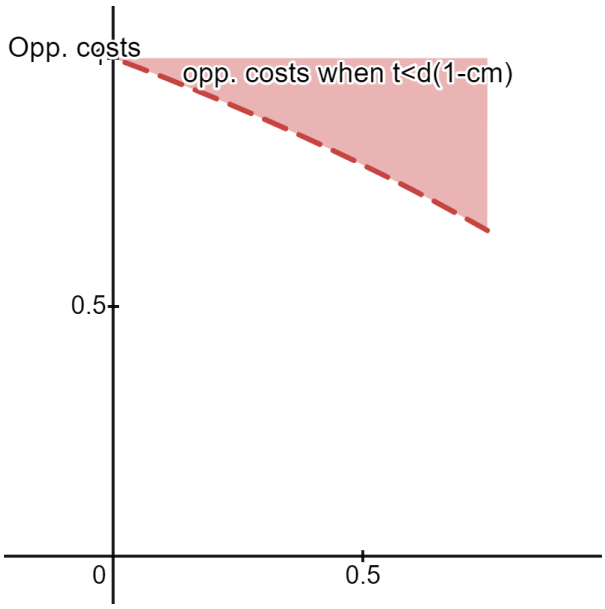
[10] Note that in the complete specification of the signaling model γ represents the opposite of opportunity costs.

[11] More concretely, in panel a, the horizontal axis is s^H (which as explained in Appendix A.4 represents the constraint for the best response vet to be better than hire) while s^F (which as explained in Appendix A.4 represents the constraint for the best response vet to be better than fire) fixed at .75. In panel a, $s^H < s^F$. In panel b, the horizontal axis is s^F, while s^H is fixed at .75 and $s^F < s^H$). Hence, In both figures, the line that is higher (s^f in Figure 6a and s^H in Figure 6b has been fixed at .75. Moreover, remaining parameters from the model in Appendix A.4 have been fixed at. $q = .75$, and $o = 1$.

Figure 6: The prevalence of pooling equilibria: In both figures, pooling equilibria in which the undesirable agent apply are in the shaded region



(a) higher costs of training. Note: horizontal axis represents how likely hiring is better than vetting.



(b) lower costs of training. Note: horizontal axis represents how likely firing is better than vetting.

costs, implying that is where semi-separation occurs. Those would be instances where the decision model is not robust. The two different models have different scope conditions attached to them. The more complex signaling model has broader scope conditions and should be used when the empirical interpretation of the theory includes regime transitions with low training costs. This however, is not the case with our empirical interpretation: post-communist Poland. As we show in the next section, our main source of data, Poland, is a clear case of weakly institutionalized regime change where the costs of training new officers were high. For this reason, in the remainder of or paper, we use the predictions associated with the pooling equilibria (reflected in the decision model) to formulate our empirical expectations.

To be clear, this is not a typical case of equilibrium selection, which are commonly dictated by theoretical reasons. Instead, theoretically, we have no reasons to believe that

pooling equilibria are more reasonable than semi-separating equilibria. The focus on them is dictated by our historical evidence. Both the focus on polling and on the best responses of the RVC strategy, (s^*) are dictated by to the empirical interpretation we are most concerned with.

4.1 Empirical implications

Our theory, while intuitive, also reveals something that was not clear before writing down the model. It is intuitive that lower costs of vetting would induce the commission to vet more. However, it is less obvious how the attractiveness of vetting compared to the alternatives - firing everyone or no one - varies with the proportion of desirable agents. Our analysis identifies an inflection point (the peak of the triangle) where the appeal of vetting shifts from increasing to decreasing with changes in cm.

Our model also has implications for what kind of purge one should expect at the wojewodztwo level, which is the level at which we have collected our data. The actionof vetting scaled up translates into a selective purge; firing scaled up translates into a thorough purge; and hiring scaled up translates into "no purge". Consequently, we infer that a selective purge is most likely at the inflection point (peak of red triangle) and its optimality decreases as one moves away from it. A thorough purge is most likely when the proportion of desirable agents is low to moderate. When the proportion of moderate and competent agents is high, "no purge" becomes optimal.

Figure 3 illustrates the propositions that can be used to produce hypotheses that we test within our paired comparison approach. To summarize, the key finding are that a selective purge becomes more attractive as the cost of screening decreases, but the expected proportion of competent and moderate agent determines which kind of purge is the alternative to a selective purge. For a low cost of selective purge, an increase in cm first increases the likelihood of selective purge (vis-a-vis thorough purge) and then decreases this probability (as "no purge" becomes more attractive).

This means that when the proportion of competent agents is low, for comparable costs of selective purges, as the proportion of extreme agents increases, a thorough purge will be chosen over a selective purge. However, when the proportion of competent agents is high, for comparable costs of carrying out selective purges, a decrease in the proportion of moderate agents will make a selective purge more attractive to "no purge".

In what follows, we will focus on illustrating the mechanisms underpinning the non-linearity implication in the empirical section using paired comparisons (Tarrow 2010). However, before we proceed to illustrate the theoretical mechanisms empirically, in the next section, we explain why data on the operation of Polish verification commissions in 1990 is ideal for our purposes.

5 Methods

To further understand the mechanisms underpinning vetting, we use the case of Poland. First, we elaborate why Poland is an ideal case to study verification of security services. Second, we select sub-units within Poland to illustrate key insights from the formal model with comparative case studies.

5.1 Poland as a case

Poland closely fits the assumptions of the model. Its democratic transition was accompanied by an abrupt economic transformation from a socialist-planned economy to a free-market one. Dubbed the "big-bang," the transition gave those with access to information ample opportunities for fast enrichment, increasing the opportunity cost of applying for security services. The political transition stretched over months.

The dual nature of regime change – from authoritarianism to democracy and from market socialism to capitalism – exacerbated the challenge that weak institutions posed to the privatization of formerly state-owned enterprises. Urban areas became hotbeds of violence

by mobsters and even soccer fans. The legacy of security forces with questionable loyalties presented an additional threat to state stability. Poland needed competent officers trained in fighting organized crime, corruption and white-collar crime.

Some agents from the former *Sluzba Bezpieczenstwa* (SB, the communist security agency) were better positioned to combat white-collar crime, ranging from corruption to industrial espionage, than newly trained recruits. Hence, the new security system was bound to include at least some personnel trained under the authoritarian SB. The question of retaining competent and moderate officers capable of staving off the multiple threats associated with the transition loomed large, especially given that in 1989, SB employed 24,308 officers (Kozlowski 2019). With a long history of repressing dissidents, it was perceived as a key pillar of authoritarianism, but any staffing policy had to consider the capacity to respond to the challenges facing the new democratic state. Poland was in dire need of a screening mechanism.

The first democratic government initially fired all officers employed by SB at the end of July 1989. Subsequently, they could apply for a position in the newly created Office for National Protection (*Urzad Ochrony Panstwa, UOP*) if they were under 55 *and* had been positively verified by the relevant RVC. Positive verification was not synonymous with re-employment in the new secret service. Out of the 14,034 security officers that underwent verification, 10,439 (or 74%) passed verification. Yet the UOP had only around 5,000 vacancies. The creation of RVCs was the result of delegating vetting decisions to the 49 wojewodztwa by the new democratic cabinet. While the decisions were decentralized, ultimately the RVCs were acting in the interest of one democratic state and their structure was uniform. Each included a UOP representative, a head of the local police appointed by the new government, a universally trusted local activist, as well as MPs and senators from the wojewodztwo. With little more than a few weeks to review hundreds of personnel files, RVC resources were thinly stretched. Given the minimal guidance from the administrative center, the variation seen in the verification outcomes from Figure 1 is hardly surprising. Some RVCs only approved officers who distinguished themselves by producing socially desirable

outcomes, while others rejected only those whose transgressions that could be proven (Kozlowski 2019). Though such idiosyncrasies are consistent with the high variation we observe in Figure 1, our theory provides a consistent reason for the variation.

Finally, recruiting and training new officers for security agencies to both supplement and replace the existing force was a difficult and time-consuming process. Attracting new recruits was particularly challenging due to the poor reputation of security agencies and relatively low pay. Recruitment was only the first hurdle; rebuilding the necessary capabilities within the security services required an estimated 8-10 years of training (2019). Although police training was shorter, taking about three years divided into basic and professional phases, this still led to a significant number of vacancies in the early 1990s (Piotrowska, Szkurłat, and Szydłowska 2024). This high cost of training reinforces our argument that, in our case, selection into screening predominantly is explained by the decision model (or followed the pooling equilibria)

5.2 Paired comparisons

In our empirical analysis we aim to demonstrate the changes in the operation of the model under different conditions. Given the size of the sample (N=49) and the emphasis on illustrating the mechanism, rather than testing the theory, a typical quantitative empirical strategy would be inadequate.

Instead, we combine quantitative and qualitative analysis, applying the model to two pairs of wojewodztwa in a paired comparison. The way we chose the cases for comparison is an adaptation of John Stuart Mill (1869)'s Method of Difference (or Most Similar Systems Design). It enables linking the key explanatory variable to the dependent variable by choosing cases that are most similar on the relevant control variables and differ on the outcome (akin in spirit to statistical matching).

In one of the few discussions concerned with combining qualitative evidence with formal models, Lorentzen, Fravel, and Paine (2017) note that most papers providing such evidence

use qualitative cases heuristically rather than systematically. This is puzzling, as the insights generated from comparative statics are often conditional on keeping the values of other variables below or above a certain threshold. This is the case in our theory, where the impact of competent agents moderates the predicted effect of extremeness. Moreover, it is worth emphasizing that in our empirical approach we are more interested in identifying the causal mechanism, rather than the causal effect of the determinants of different purge types (Gerring 2004).

Critics of the Empirical Implications of Theoretical Models (EITM) program point out that even though a formal model isolates the effects of parameters omitted from the formal model, one cannot do the same when carrying the model's parameters over to a regression framework. There, any omitted variables can bias the results. However, the paired comparison approach is particularly suitable for examining the implications of comparative statics in a formal model because it allows us to control for the model's parameters by using parameters of the model to select cases (Goemans and Spaniel 2016).

Moreover, a sub-national study allows us to control for broader country-specific variables. Given that the 49 RVCs operated independently, this is reminiscent of a sub-national analysis of, say, state legislatures in the 50 United States. This is a crucial advantage in light of Tarrow (2010)'s criticism of the use of this method by Skocpol (1979) and Putnam (1994). Tarrow argued that neither of these authors accounted for hidden cultural or historical determinants. In our analysis the differences and similarities are all the result of choice, rather than happenstance, however all the units are a part of the same country and were subject to the same overarching secret service reform.

While paired comparisons have been identified as a useful tool for illustrating theoretical arguments, we are so far short of a method that would allow us to systematically pick the best, indeed most similar cases. The existing approaches to marrying formal theory and qualitative approaches often concentrate on "important" cases (e.g., the analytic narratives approach describes the case selection as "our cases selected us, rather than the other way

round" (Bates 1998)) or, in the case of paired comparisons, rely on the readers' trust that the cases are indeed most similar.

Hence, the main methodological innovation of this paper is how we identify our cases for analysis. Our contribution is to introduce a systematic method for choosing the cases for the comparative case study using a fairly data-intensive quantitative approach.

We proceed as follows. First, we identify pairs of wojewodztwa with different purge types that are geographical neighbors and who were part of the same historical partition[12]. A list of such pairs is provided in Appendix B.6. For each pair, we calculate the Euclidean distance, δ, between the values of the key variables. The Euclidean distance between wojewodztwa 1 and 2 described by two variables (x, and y), and coordinates (x_1, y_1) and (x_2, y_2) is calculated as:

$$\delta = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{3}$$

x and y, correspond to the values of population density and competence for a given wojewodztwo. This allows us to identify the pairs with the smallest distance, i.e. the most similar cases. As mentioned above, this method is relatively data-intensive: to identify the perfect cases, we needed to collect wojewodztwo-level data on multiple relevant variables, only to then discard all but our four chosen cases. However, it allows us to make a convincing case that a pair contains units that are the *most* similar.

5.3 Data

Before we conduct the paired comparisons to showcase the causal mechanisms present in the formal model, we need to operationalize its key variables: purge types, cost of RVC operation, extremeness, and competence.

[12] This acknowledges legacies of the fact that for 123 years Poland's territory was partitioned by three empires: Russia, Prussia, and Hapsburg.

5.3.1 Purge types

To categorize each wojewodztwo, we analyzed several historical and archival sources, compiling them into an original dataset. As an illustration, we provided a page from the files we coded (a transcript from RVC proceedings) in Figure B3 in Appendix B.3. Files of the 49 RVCs allowed us to measure the key variables of interest as they contained, among others, information on:

1. the number of verified officers;

2. the number of those verified positively;

3. the time each RVC spent on deliberation.

To classify purges according to our theoretical definitions, we need to establish their severity *and* how carefully the RVC considered each application. Selective purges are those where commissions read files closely. Whether this process led to many or few negative decisions was determined by candidate quality. Conversely, in thorough and "no purges" decisions were taken swiftly, the difference between the two lied in whether many or few applications were rejected.

Hence, a *selective purge* is one where the RVC spent considerable time deliberating each case; a *thorough purge* is one where RVCs took little time deliberating *and* verified a high proportion of officers negatively; and *no purge* is where the RVCs took little time deliberating *and* cleared a high proportion of officers to serve under the new democratic system. Thus, *no purge* means that the commission was relatively lenient, not necessarily that all the agents were verified positively. Similarly, a *thorough purge* means that the commission was relatively stringent. Because our empirical equivalents are more flexible than the theoretical purges, to distinguish them, henceforth we use italicized names to indicate the empirical purge classifications.

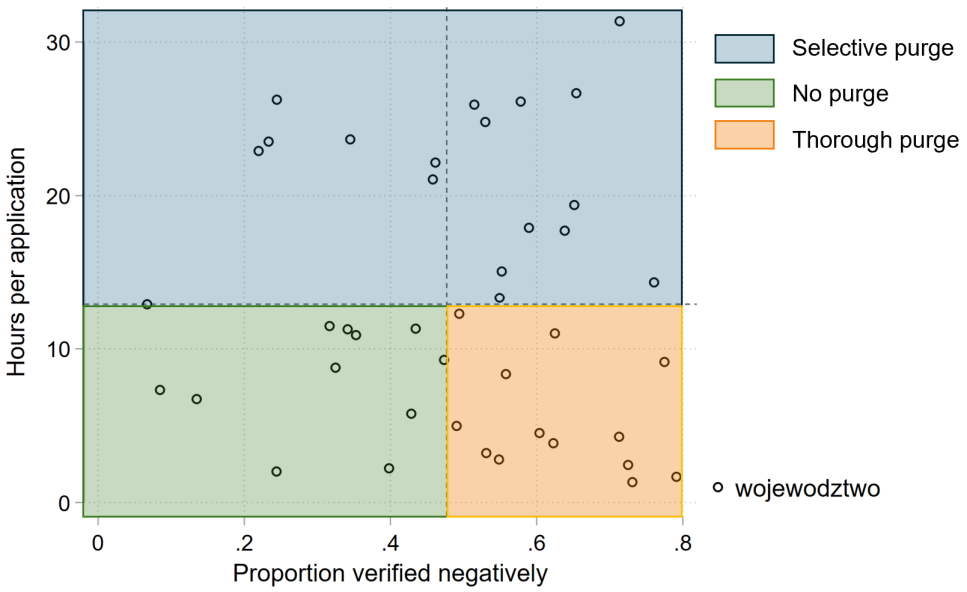
Translating "less" or "more" time into clear-cut categories requires cutoffs. In Table 1 we

construct such cutoffs using the mean share of negatively verified [13] and the mean number of hours expended per file[14].

Table 1: Types of purges

Purge type	Proportion verified negatively	Hours per application	Frequency	Percent
No	<0.48	<12.9	12	28.57
Selective		>12.9	17	40.48
Thorough	>0.48	<12.9	13	30.95

Figure 7: Classifying purge types



This rule allows us to classify purge types in the majority of wojewodztwa.[15] Figure 7 depicts all wojewodztwa for which we have the information and places them in a space defined by the proportion verified negatively and the time spent per application. The lines demar-

[13] That it, the number of those whose verification was negative/Number of applicants in a given wojewodztwo

[14] (That it, the total time spent deliberating times number of commission members)/Number of applications. The time spent deliberating is calculated as the time between the beginning and end of RVC operation

[15] Some wojewodztwa could not be classified as the dates of operation of the RVCs were missing from the files.

cating the three types of purges are set to the mean values of the two variables. Appendix B.4.1 contains archival evidence (letters to the RVC) and data on appeals to support the robustness of our classification. In the same appendix, we present two alternative categorizations: one using variable medians, rather than means (Table B2), and one using a K-means algorithm (unsupervised classification algorithm) to identify the three categories. Comparing the three approaches, the categorization is relatively robust to the cutoff definition (Table **??**). Crucially, the four cases used for the paired comparison are not affected.

5.3.2 Moderation, competence, and the cost of running an RVC

The key explanatory variables in the mechanism of verification are: the level of competence among officers in a given wojewodztwo, c; the extent of moderation among the officers, m; and the cost of conducting a selective purge, s. To operationalize them, we build on data from the IPN and historical statistical yearbooks.

First, to measure the competence of the officer corps of the SB, we use data from the IPN on the distribution of officers working across various departments, disaggregated by region (Piotrowski 2003). The proportion of officers employed pre-transition in the Department for the Protection of the Economy (PoE, *Wydział Ochrony Gospodarki*) in 1989 is particularly relevant. Since it was tasked with securing economic resources, the prestige of the department attracted highly capable officers prior to the transition (Kozlowski 2019). Moreover, their skills can be interpreted as "usable" (Grzymala-Busse 2002) by the new democratic regime because, in contrast to officers who focused on repressing the anticommunist opposition or the Catholic Church, officers from PoE were trained to solve crimes that were economic rather than political. Hence, we use the proportion employed in PoE as a proxy for competence (Reported in Table 2). As an alternative measure, we use the data from an online IPN catalog including a sample of SB functionaries to measure the proportion of the officers who enrolled in additional courses or training while in service (described in Appendix B.4.3). While these courses might not have included skills relevant to democratic policing, they are

a costly signal of commitment to skill-building.

Second, to capture extremeness (that is the reverse of m from the theoretical model) among officers, we use the intensity of repression during Martial Law (1981-1983), measured as the number of persons convicted in the immediate aftermath of the announcement of Martial Law. All the cases we include were political, as political offenses were criminalized according to Martial Law provisions. To construct our first measure, we divide the number of convictions by the number of people that were brought to court (Instytut Pamieci Narodowej 2021) in each wojewodztwo. On average, 63.9% of those tried were sentenced. Although courts and security services were officially separate organizations, research by Popova (2012) and McCarthy (2015) suggests that in autocracies, the actions of these institutions are more rigidly aligned than in democracies. In Poland, these institutions were fused, acting in sync to protect the interests of the PZPR party. Hence, we feel confident using repression by courts as a proxy for extremism among local security services.

We acknowledge that measuring repression is bogged by problems of reverse causality (Ritter and Conrad 2016). Recognizing this reciprocal relationship, for our main measure of extremeness, we normalize the court sentences with the number of Solidarity members in each wojewodztwo to control for the level of dissent. We also used an alternative measure - the number of murders or disappearances by SB in each wojewodztwo during the 1980s — but rejected it as the primary measure due to its high threshold of repression. In 18 of the 49 wojewodztwa, no murders by the secret police were confirmed. However, we still use it to support the claim that our chosen pairs of wojewodztwa are the most similar (See Appendix B.4.3).

To account for the cost of running RVCs, we use 1989 population density data from the Statistical Yearbook of Poland. The costs of selective purges were linked to the salience of verification. In Poland, more urban wojewodztwa with fewer rural areas had a denser network of dissident activists, as Solidarity initially organized in large workplaces. Consequently, urban areas had a higher likelihood of residents personally knowing someone persecuted

by the secret police, making the conflict between citizens and the SB more prominent. During verification, this heightened the stakes for both the opposition and the former security services, leading both sides to pressure the RVC to rule in their favor. In contrast, lower density areas saw less pronounced conflict, reduced pressure from both groups, and allowed the commission to thoroughly review the files. The Minister of Interior at the time, Krzysztof Kozlowski, argued that while in areas where people knew each other, the RVCs experienced pressures for revenge, in places with weaker networks, the process was more systematic and devoid of emotions (Kozlowski 2019, p.206).[16]

Building on these insights, we assume that the smaller a wojewodztwo's population density, the lower the RVC's cost of engaging in a scrupulous examination of individual officer files. In the robustness checks (Appendix B.4.3), we consider two alternative operationalizations of commission operation cost: wojewodztwo-specific average salaries and wojewodztwo revenues per capita. Both are meant to capture the opportunity costs of participating in an RVC by its members (see Table B3). At the same time, more densely populated areas would have more SB officers. Consequently, the time any given RVC member could spend on a file decreased with population density, increasing the cost of a selective purge also.

We do not include t (cost of training replacement) in our empirical approach, as it is is likely to be common for the entire country, due to the centralized nature of security service training [17] and, consequently, does not enter directly into our empirical analysis.

Summarizing, the preferred operationalizations of the key parameters are:

- c - the proportion of competent agents - the proportion of officers who worked in the Department for the Protection of the Economy;

- s - the cost of setting up the commission - population density;

[16] We acknowledge that this could seem to make "no purge" equally costly, but RVCs could reach "no purge" inconspicuously, without interviewing former secret police agents or soliciting input from the community and drawing attention to the RVC. Hence, the combination of visible effort on part of the commissioners coupled with the decision to positively verify some of them is what we believe is costly.

[17] At the time when RVC's were operating, the main campus for training secret police officers was in Legionowo, a suburb of Warsaw.

Table 2: Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Verified negatively	49	0.477	0.198	0.067	0.791
Hours/application	42	12.90	8.64	1.32	31.35
Population ('000)	49	776.498	599.4678	246.2	3968.3
Population density	49	0.143	0.146	0.0444	0.749
Repression (cases)	48	65.71	100.97	0	564
Repression (sentences)	49	36.7	50.29	0	213
Repression (murders)	49	1.96	3.21	0	14
Solidarity membership	49	196331.9	187110.9	33853	1122114
Competence (PoE)	48	0.27	0.09	0.13	0.53
Competence (courses)	49	0.42	0.14	0.12	0.75

- m - the probability that the agent is moderate - 1-(the proportion repressed during the Martial Law, standardized by Solidarity membership).

We use these variables to calculate the distances δ between the wojewodztwa. In Appendix B.4.3, we show that alternative specifications of the variables show a similar relative ordering of the pairs along the relevant dimensions.

5.4 Analysis

We focus the empirical analysis on illustrating the the non-linearity of the likelihood of observing a selective purge in cm: For low levels of cm, as the desirability of agents increases, selective purges become more likely (compared to thorough purges). Upon reaching the inflection point (for higher levels of cm), further increases in desirability lead to a decrease in the likelihood of a selective purge (as "no purge" becomes more likely).

5.5 Paired comparisons

We use the MSSD approach outlined above to identify the two most similar pairs of wojewodztwa in terms of population and competence. These happen to also correspond to

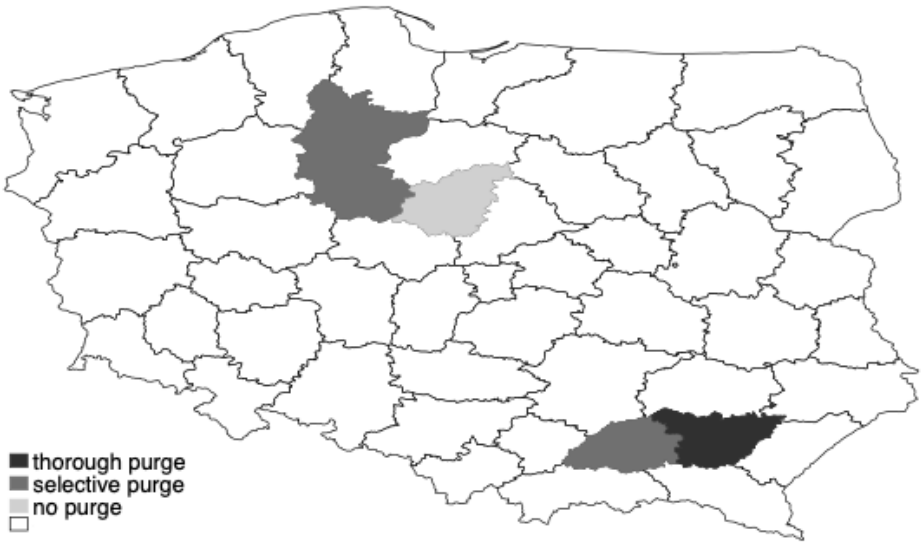
choices between a selective purge and the two purge alternatives, and span a wide range of competence and moderation (See Table 3 and Figure 8).

Table 3: wojewodztwa compared and their characteristics

Wojewodztwo	Repression	Population density	Competence	Purge type	Officers verified
Tarnowskie	0.0000705	0.160	0.211	selective	139
Rzeszowskie	0.0002542	0.163	0.213	thorough	206
Bydgoskie	0.0002925	0.107	0.363	selective	284
Wloclawskie	0.0000778	0.097	0.354	no	164

The wojewodztwa satisfy a Most Similar Systems Design: the two pairs share similar levels of competence (agents in PoE), cost (population density), historical and cultural legacies, and have in common a host of other variables, as discussed below. At the same time, one wojewodztwo in each pair saw a *selective purge* while the other saw a *thorough* or *no purge*, respectively. This allows us to control for common systemic characteristics while inter-systemic differences play the role of explanatory variables, making the wojewodztwa counterfactuals for each other (Przeworski and Teune 1981; Goemans and Spaniel 2016).

Figure 8: Location of the compared wojewodztwa



Consider first the two wojewodztwa with similarly low levels of PoE agents (competence): Tarnowskie and Rzeszowskie (Figure 8). They experienced different types of purges even though they are very similar even beyond the key explanatory variables. First, their share of city and village districts (gminas) is nearly identical, signifying a similar urban-to-rural composition. Second, they both lie in a part of Poland that from the 18th century was controlled by the Hapsburg Empire, leading to shared institutional legacies. Third, the size and composition of their RVCs was similar: they were rather small (eight members each), and both were headed by prominent Solidarity activists.

Rzeszowskie underwent a *thorough purge*, with all 206 applications considered in three days and 54% verified negatively. Yet, the purge in Tarnowskie was *selective* – it took 19 days to reject 39% of the 139 applications. The stringency of the decision was highlighted by the collective resignation of members of the Tarnowskie commission in response to the introduction of an appeals process (Wszolek 2019).

The main difference between the two wojewodztwa was the extremism of the agents. Rzeszowskie experienced brutal repression throughout the 1980s. In the early days of Martial Law, around 100 activists were arrested and 163 more were threatened. In the following years, security forces violently broke up anti-communist demonstrations. For example, a 1982 demonstration resulted in 136 arrests (including 42 students) and the hospitalization of five individuals (Gliwa, n.d.). Repression also claimed the life of one dissident (Instytut Pamieci Narodowej 2022). Further evidence of how ideologically extreme Rzeszowskie's secret police is that following the announcement of the 1989 elections (won by Solidarity) they were ready to assist the Ministry of Interior in invalidating those results and imposing a second version of Martial Law. IPN archives contain lists of oppositionists that had been drawn up by the secret police in Rzeszowskie to carry out "warning conversations" and another list of individuals to be arrested (Draus and Nawrocki 2000).

In contrast, in Tarnowskie, the number of activists arrested and placed in internment camps was relatively low (Redakcja 2006) and included no fatal casualties. According to the

memoirs of Solidarity leaders in the region, any persecution of the trade union was the result of exceptionally eager informants entangled in Solidarity leadership (Leslaw Maleszka and Zbigniew Stanuch) rather than a particularly repressive local security agencies. Documents of the Tarnowskie RVC show evidence that a relative lack of public pressure lead to a careful consideration of cases: "(i)n cases where the in-depth information did not give unequivocally negative indications, the RVC requested positive consideration of the appeals" (BU/3546/48 p.8).

Hence, while both wojewodztwa had similarly low levels of competence and similar population density (s), Rzeszowskie was lower on the cm dimension due to high extremism, making the thorough purge more likely, as shown in Figure 9. The figure places Rzeszowskie and Tarnowskie in the context of our model, using the narrative above and our operationalization of model parameters and illustrates our theoretical results from section 3.

We conduct a similar exercise for Wloclawskie and Bydgoskie, our second pair of wojewodztwa, which are characterized by relatively high competence: the proportion of agents employed in PoE was 36% and 35%, respectively.

Beyond the variables operationalizing our model's parameters, Bydgoskie and Wloclawskie shared institutional legacies: they remained under Prussian control for a significant part of the Partition period. In fact, until the 1975 administrative reform, they were assigned to the same administrative unit. Bydgoskie and Wloclawskie also had comparable levels of "Solidarity" membership: 27% in Bydgoskie and 21% in Wloclawskie. Finally, they had large RVCs, with 17 and 13 members respectively. The two wojewodztwa differed, however, in their extremism —our explanatory variable—and the type of purges they experienced—our dependent variable.

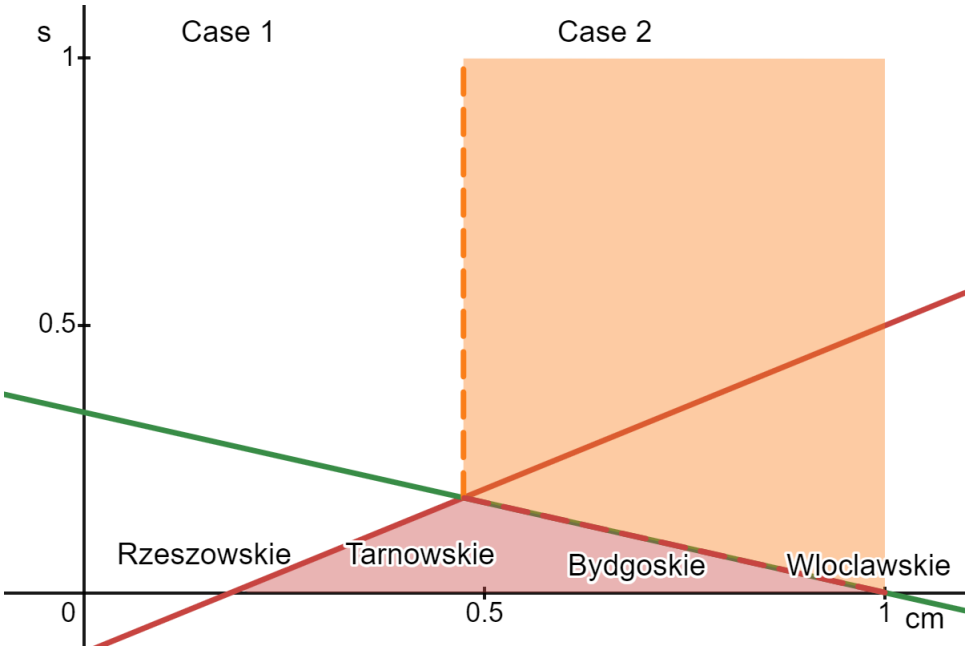


Figure 9: Tarnowskie vs Rzeszowskie and Wloclawskie vs Bydgoskie in light of model predictions (orange - no purge, red- selective, white - thorough purge)

The RVC in Bydgoskie chose a *selective purge*. The commission members spent 20 days carefully vetting the files of SB. The commission noted a "considerable public approval of the work of the RVC. Demands for revenge, if there were any, soon morphed into acceptance and even empathy" (By/453/47 p.204). This seemed surprising given the high level of repression in Bydgoskie especially before the introduction of Martial Law in 1981. Violence culminated with severe beatings of Solidarity members (Chincinski 2002). The event led to unrest throughout Poland but particularly in Bydgoskie.

Conversely, in Wloclawskie, the commission chose *no purge*, letting through 66% applicants in seven days. The outcome sparked protests from opposition members who argued to the Ministry of Interior that a high level of positive verification was morally wrong, given the oppression Solidarity suffered under the SB (BU/3546/52 p.63-69). This suggests pressure from Solidarity affiliates, which is consistent with mechanisms we label as *no purge*.

We attribute the difference in purge types between the two wojewodztwa to the difference in agents' extremism, as posited by our model. The higher level of extreme agents in Byd-

goskie than in Wloclawskie, while holding constant the high level of competence, according to our theory, made a selective purge comparatively more likely in Bydgoskie, as it lowered *cm*. Wloclawskie experienced fewer judicial persecutions and had no reported SB murders or disappearances, unlike Bydgoskie, which reported four cases.

Our model thus helps us understand a potentially puzzling phenomenon: the relative leniency of the RVC in Bydgoskie (*selective purge*, 35% negative verification rate), compared to Rzeszowskie (*thorough purge* and 55% negative verification), despite much higher levels of repression in the former. We argue that this is caused by two mechanisms. First, the suppressed public desire for revenge gave RVC in Bydgoskie latitude to choose a selective purge. Second, the relatively high competence of security officers made a thorough purge unattractive. Hence, our model helps understand why retribution, which seems to be a natural response to repression, gave way to practicality.

6 Scope conditions

Our model is applicable more broadly to post-authoritarian countries under two conditions: the state retained control over the files of their former secret police and the secret police was a relatively centralized and powerful agency. The first condition places control over personnel files in the hands of the democratic government. The rationale behind the second condition is based on the summarized earlier work by Sheena Chestnut Greitens (2016), according to whom secret police centralization depends on the origin of threats to authoritarian rule. Though insightful at a macro level, Chestnut Greitens' argument does not delve into how the different secret police agencies ought to be reformed should the regime undergo a democratic transition. Our model applies specifically to those new democracies that follow autocracies where the dictator feared revolution from below, resulting in a centralized security service. While the post-communist countries of Eastern Europe, including Poland, share this characteristic, so do other autocracies (Spain under Franco, Mexico under PRI in Latin America,

and, arguably, China under Xi Jinping). In Appendix C, we analyze the case of Ukraine for an out-of-sample test.

Post-Soviet states of Eastern Europe with their powerful secret police agencies easily satisfy the centralization requirement. However, not all authoritarian regimes delegate such broad powers to secret police agencies. Whereas we would apply our model to these non-Eastern European states with caution, we feel quite confident in its relevance for post-communist Eastern Europe and believe that Poland is an ideal representative of those states. Poland shares the dual nature of regime change with post-communist countries, including transitions from authoritarianism and from a planned economy. It also shares with many the gradual and the negotiated features of transition process. The slow pace of transition could factor into the price of selective purges, making careful verification cheaper as the files of the secret police could be secured through a political rather than revolutionary transition. Kozlowski (2019) documents such destruction in the multi-week-long transition period.

In our case, the high cost t of training new personnel influenced both the likelihood of different purge types and reduced the severity of concerns related to self-selecting into the screening process. In other contexts, training costs might be lower, impacting self-screening dynamics and the desirability of selective purges vis a vis "no" or thorough purges. Lower training costs are more likely in countries undergoing abrupt transitions, particularly if the transformation does not include the old regime's security services in negotiating the transition. In such cases, recruits may be easier to enlist, as the new police force is less burdened by the old regime's reputation. The rapid pace of change may also lead to shorter, simplified training programs due to heightened momentum. These factors can reduce the cost of training new personnel, affecting the empirical implications of the model.

Taken together, this suggests that the type of purges and self-selection problems could be different in countries where transition was more abrupt. Nevertheless, the underlying problem modeled here – the combination of personnel shortages with acute uncertainty concerning competence and loyalty of state employees to the former authorities – is ubiquitous.

7 Conclusion

Post-authoritarian governments face a dilemma: how to reform or replace compromised security agencies without undermining essential state functions like the protection of life, liberty, property, and contract. New democracies must decide how to handle personnel from these agencies. The recommendation to simply remove "bad apples"—is inconclusive, as it could refer to dismissing incompetent officers ideologically loyal officers, or both. Moreover, competent and neutral officers could be valuable to the new regime but are not easily identifiable. We argue that in contrast to authoritarian purges which could be more or less severe, new democratic governments have available to them three options that cannot be arranged on a single dimension: a thorough purge, "no purge", or a "selective purge. " The latter refers to carefully vetting and retaining as a result of the vetting desirable agents. We show that the even though loyalty and competence feature prominently in decisions of vetting commissions, instead of setting up a trade-off between loyalty and competence, the two features characterize the selection problem of commissions and require a model specific to regime transitions.

Our article thus explores the conditions under which selective purges are chosen and when other approaches might prevail. It offers both theoretical and empirical contributions.

On the theoretical front, we present a simple framework that situates familiar concepts, such as the loyalty-competence dilemma, within the context of regime change to identify when new democracies should carefully verify their security agencies through a process we term a "selective purge." We show that selective purges are more likely when verification costs are low. Additionally, we illustrate how the choice of purge type depends on the average competence and loyalty among security officers. Our findings indicate that when competence and moderation (cm) are high, a selective purge is unnecessary, as most agents are already suitable; therefore, "no purge" is preferred unless the costs of training new staff are very low. However, as cm decreases (when either competence or moderation falls), the advantage of "no purge" diminishes. If screening costs (s) are high, a thorough purge becomes more

attractive as *cm* declines. Conversely, if screening costs are low, a selective purge becomes more favorable for both low and high *cm* values. In summary, the model's predictions depend on the proportion of competent and moderate agents: if it is low, the choice is between a thorough and selective purge; if it is high, the choice is between a selective purge and "no purge."

Furthermore, a selection model supporting the assumptions of the above analysis challenges the conventional wisdom regarding verification commissions. This conventional view suggests that the mere threat of a commission uncovering the truth in personnel files deters undesirable officers from seeking verification. However, this assumption relies on the existence of a separating equilibrium that is, in practice, unattainable. Therefore, the existence of verification commissions alone does not fully deter low-quality applicants.

Our theoretical findings are supported by a comparative analysis using data from the operation of 49 regional verification commissions in Poland in 1990. The paired comparison approach we adopt is a novel application of the Most Similar Systems Design to illustrating a formal model. To identify cases that are indeed most similar, we propose using the Euclidean distance calculation. The two pairs of wojewodztwa identified this way allows us to show how, for otherwise nearly identical units, the higher level of repression decreases (for low average quality of applicants) and increases (for high quality) the likelihood of a selective purge.

Bibliography

Banks, Jeffrey. *Signalling games in political science*. Routledge, 2013.

Bates, Robert H. *Analytic narratives*. Princeton University Press, 1998.

Blaydes, Lisa. *State of Repression: Iraq under Saddam Hussein*. Princeton University Press, 2018.

Bokobza, Laure, Suthan Krishnarajan, Jacob Nyrup, Casper Sakstrup, and Lasse Aaskoven. "The morning after: cabinet instability and the purging of ministers after failed coup attempts in autocracies." *The Journal of Politics* 84, no. 3 (2022): 1437–1452.

Callander, Steven, et al. "A theory of policy expertise." *Quarterly Journal of Political Science* 3, no. 2 (2008): 123–140.

Chincinski, Tomasz. "Bydgoski marzec 1981 roku." *Biuletyn Instytutu Pamieci Narodowej* 2, nos. 12 (23) (2002).

Clarke, Kevin A, and David M Primo. *A model discipline: Political science and the logic of representations*. Oxford University Press, USA, 2012.

Crabtree, Charles, Holger L Kern, and David A Siegel. "Cults of personality, preference falsification, and the dictator's dilemma." *Journal of Theoretical Politics* 32, no. 3 (2020): 409–434.

David, Roman. "Lustration laws in action: The motives and evaluation of lustration policy in the Czech Republic and Poland (1989–2001)." *Law & Social Inquiry* 28, no. 2 (2003): 387–439.

Draus, Jan, and Zbigniew Nawrocki. Przeciw Solidarnosci: Rzeszowskia Opozycja w Tajnych Archiwach Ministerstwa Spraw Wewnetrznych.

Dudek, Antoni, and Andrzej Paczkowski. "Poland." Edited by Krzysztof Persak and Łukasz Kaminski. In *A Handbook of the Communist Security Apparatus in East Central Europe 1944-1989*. (No Publisher), 2005.

Egorov, Georgy, and Konstantin Sonin. "Dictators and their viziers: Endogenizing the loyalty–competence trade-off." *Journal of the European Economic Association* 9, no. 5 (2011): 903–930.

Ekiert, Grzegorz. *The state against society: Political crises and their aftermath in East Central Europe*. Princeton University Press, 1996.

Epstein, David, and Sharyn O'Halloran. *Delegating powers: A transaction cost politics approach to policy making under separate powers*. Cambridge University Press, 1999.

Fudenberg, Drew, and Jean Tirole. *Game theory*. MIT press, 1991.

Gehlbach, Scott. *Formal models of domestic politics*. Cambridge University Press, 2021.

Gerring, John. *Case study research: Principles and practices*. Cambridge university press, 2016.

———. "What is a case study and what is it good for?" *American political science review* 98, no. 2 (2004): 341–354.

Gibbons, Robert S. *Game theory for applied economists*. Princeton University Press, 1992.

Gliwa, Malgorzata. Region Rzeszowski NSZZ S.

Goemans, Hein, and William Spaniel. "Multimethod research: A case for formal theory." *Security Studies* 25, no. 1 (2016): 25–33.

Goldring, Edward, and Austin S Matthews. "To purge or not to purge? An individual-level quantitative analysis of elite purges in dictatorships." *British Journal of Political Science* 53, no. 2 (2023): 575–593.

Greitens, Sheena Chestnut. *Dictators and their secret police: Coercive institutions and state violence*. Cambridge University Press, 2016.

Grzymala-Busse, Anna M. *Redeeming the communist past: The regeneration of communist parties in East Central Europe*. Cambridge University Press, 2002.

Hassan, Mai. *Regime threats and state solutions: Bureaucratic loyalty and embeddedness in Kenya*. Cambridge University Press, 2020.

———. "The strategic shuffle: Ethnic geography, the internal security apparatus, and elections in Kenya." *American Journal of Political Science* 61, no. 2 (2017): 382–395.

Hassan, Mai, Daniel Mattingly, and Elizabeth R Nugent. "Political control." *Annual Review of Political Science* 25, no. 1 (2022): 155–174.

Horne, Cynthia M. *Building trust and democracy: Transitional justice in post-communist countries*. Oxford University Press, 2017.

Huber, John D, and Charles R Shipan. *Deliberate discretion?: The institutional foundations of bureaucratic autonomy*. Cambridge University Press, 2002.

Instytut Pamieci Narodowej, . Elektroniczny inwentarz archiwalny akt spraw karnych z okresu stanu wojennego.

———. "Ofiary Stanu Wojennego – niemal sto smierci W niewyjaśnionych okolicznosciach." *Instytut Pamieci Narodowej - Poznan*, 2022.

Interview. Anonymised Structured Interview with Authors. Interview.

Ketchley, Neil, and Gilad Wenig. "Purging to Transform the PostColonial State: Evidence from the 1952 Egyptian Revolution." *Comparative Political Studies*, 2023.

Komar, Andrzej B, and Marian Cz. Niedzialek. "Przyczyny Zwolnien z Resortu Spraw Wewnetrznych." *Humanizacja Pracy* 6, no. 1 (1990): 40–52.

Kozlowski, Tomasz. *Koniec imperium MSW: transformacja organow bezpieczenstwa panstwa 1989-1990*. Instytut Pamieci Narodowej–Komisja Scigania Zbrodni przeciwko Narodowi, 2019.

Lorentzen, Peter, M Taylor Fravel, and Jack Paine. "Qualitative investigation of theoretical models: the value of process tracing." *Journal of Theoretical Politics* 29, no. 3 (2017): 467–491.

Mack, Eric. "Robert Nozick's Political Philosophy." Edited by Edward N. Zalta. In *The Stanford Encyclopedia of Philosophy*, Summer 2018. Metaphysics Research Lab, Stanford University, 2018.

Marat, Erica. *The Politics of Police Reform: Society Against the State in Post-Soviet Countries*. Oxford University Press, 2018.

Mattingly, Daniel C. "How the party commands the gun: The foreign–domestic threat dilemma in China." *American Journal of Political Science* 68, no. 1 (2024): 227–242.

McCarthy, Lauren A. *Trafficking Justice: How Russian Police Enforce New Laws, from Crime to Courtroom*. Cornell University Press, 2015.

Mill, John Stuart. *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Princilples of Evidence and the Methods of Scientific Investigation*. Harper; brothers, 1869.

Montagnes, B Pablo, and Stephane Wolton. "Mass purges: Top-down accountability in autocracy." *American Political Science Review* 113, no. 4 (2019): 1045–1059.

Nalepa, Monika. *After Authoritarianism: Transitional Justice and Democratic Stability*. Political Economy of Institutions and Decisions. Cambridge University Press, 2022.

Nozick, Robert. *Anarchy state and utopia*. London, England: Blackwell, 1978.

Oseka, Piotr. "Funkcjonariusze SB 1970-1989—drogi awansu i modele kariery." *Wykluczeni*, 2008, 108–124.

Piotrowska, Barbara Maria. "The price of collaboration: how authoritarian states retain control." *Comparative Political Studies* 53, no. 13 (2020): 2091–2117.

Piotrowska, Barbara Maria, Izabela Szkurłat, and Magdalena Szydłowska. "Regime transitions and institutional weakness: the case of police reform in Poland in the early 1990s." *Street-Level Bureaucracy in Weak State Institutions*, 2024, 137.

Piotrowski, Pawel. "Struktury Sluzby Bezpieczenstwa MSW 1975-1990." *Pamiec i Sprawiedliwosc* 3, no. 1 (2003): 51–108.

Popova, Maria. *Politicized justice in emerging democracies: a study of courts in Russia and Ukraine*. Cambridge University Press, 2012.

Przeworski, Adam, and Henry Teune. "The Logic of Comparative Social Inquiry, rev. ed." *NY: Wiley & Sons*, 1981.

Putnam, Robert D, Robert Leonardi, and Raffaella Y Nanetti. "Making democracy work." In *Making democracy work*. Princeton university press, 1994.

Redakcja. Wspomnienia Po Latach.

Ritter, Emily Hencken, and Courtenay R Conrad. "Preventing and responding to dissent: The observational challenges of explaining strategic repression." *American Political Science Review* 110, no. 1 (2016): 85–99.

Scharpf, Adam, and Christian Glassel. "Why underachievers dominate secret police organizations: evidence from autocratic Argentina." *American Journal of Political Science* 64, no. 4 (2020): 791–806.

Skocpol, Theda. *States and social revolutions: A comparative analysis of France, Russia and China*. Cambridge University Press, 1979.

Spence, Michael. "Job market signaling." In *Uncertainty in economics*, 281–306. Elsevier, 1978.

Sudduth, Jun Koga. "Coup risk, coup-proofing and leader survival." *Journal of Peace Research* 54, no. 1 (2017): 3–15.

Svolik, Milan W. *The politics of authoritarian rule*. Cambridge University Press, 2012.

Tarrow, Sidney. "The strategy of paired comparison: toward a theory of practice." *Comparative Political Studies* 43, no. 2 (2010): 230–259.

Thomson, Henry. *Watching the Watchers: Communist Elites, the Secret Police and Social Order in Cold War Europe*. Cambridge University Press, 2024.

Wszolek, Grzegorz. *Sluzba Bezpieczenstwa w Krakowie na Tle Przemian w Ministerstwie Spraw Wewnetrznych 1989-1990*. Instytut Pamieci Narodowej, 2019.

Zakharov, Alexei V. "The loyalty-competence trade-off in dictatorships and outside options for subordinates." *The Journal of Politics* 78, no. 2 (2016): 457–466.

Appendix

A Formal Appendix

This appendix provides supplementary information and solutions pertaining to the signaling model described in the main text.

A.1 Robustness check for the decision model: the signaling model

Recall that to debunk the conventional wisdom on how verification commissions work and additionally check the robustness of our decision model, we propose a signaling model where an agent with private information decides whether to apply to be verified. Failure to apply, regardless of type, results in an outside option, o, representing opportunities for employment in the young economy, which are abundant for anyone, regardless of qualifications and political orientation.[18] This accounting for opportunity costs of applying in an extension relative to the original baseline model. As in the baseline, the RVC knows only the aggregate distribution of competence and loyalty in its district. However, it can pay a cost, s, to learn the loyalty and competence of the the agent. As in the baseline model, this is the cost of taking the time to examine his personnel file. Such files contain information about their performance: whether they received awards or sanctions, their tenure on the job, career trajectory, transfer history and education acquired before and during their service. The agent does not know how costly reading and studying files is to the RVC. If the RVC decides to vet the agent, then there is an exogenous probability $q \in (0, 1)$ that evidence of the officer's incompetence or extremeness survived the transition. Hence, with $1 - q$ even if the RVC embarks on a selective purge, the vetting may not turn up evidence of extremeness or incompetence of undesirable types. This is a departure from the baseline model that is motivated not so much by the need for a robustness check, but to add realism to our model. In Appendix B.2 we have assembled an abundance of secondary sources describing the burning and shredding of files at the time of the transition that was difficult for us to ignore when constructing this more detailed model.

We build on the analysis in section 2 and assume that, since the RVC only wants to retain the moderate and competent type of agent, we can reduce the type space to just two. We will refer to the moderate and competent type of agent as the **desirable** type and call the other three types simply **undesirable**. The informational assumptions of the model are:

1. The RVC, in contrast to the agent, knows its cost of vetting;
2. At the start of the game, the RVC knows only the overall distribution of desirable and undesirable agents but not who among those who applied for verification is desirable or undesirable;
3. The officers know if their type is desirable or undesirable, but do not know the cost of vetting of the specific RVC that is able to purge them. They assume the cost is uniformly distributed between 0 and 1.
4. No one knows if evidence against the undesirable types survived the transition.

[18] For instance, extreme officers can always find employment in private security, incompetent ones can find jobs in new firms that are being created (Wszolek 2019).

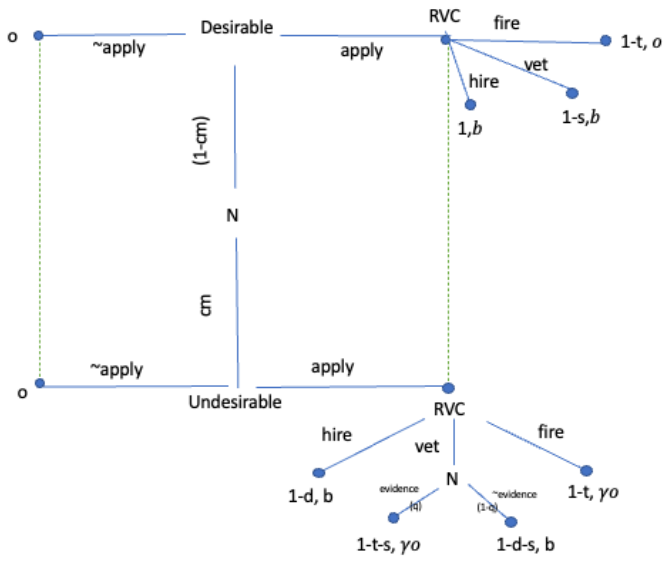


Figure A1: The Verification Game

Given these assumptions, we are interested in what kind of purge the RVC chooses and when do officers of the former secret police apply for verification? Specifically, we want to know whether undesirable agents apply at the same rate as desirable agents.

The game tree representing the interaction between the RVC and agent is presented in Figure A1 with the timing of the game and payoffs provided below the game tree.

1. Nature selects the type of officer: desirable (with with probability cm) vs undesirable (with probability $1 - cm$).

2. The officer observes his type and decides to apply or not to the RVC for verification. If he does not apply, the game ends.

3. RVC selects chooses between hiring, firing and vetting

4. If hire or a fire is chosen, the game ends and payoffs are distributed.

5. If the RVC chose vetting, Nature reveals evidence of competence and moderation, if available, with probability q (with probability $1 - q$ there is no evidence against undesirable agents). If evidence against undesirable types exists, they are passed over, otherwise officers are reappointed. All remaining payoffs are distributed.

We assume $1 > d > t > 0$, as well as $\gamma < 1 < o < b$.

The interaction between secret police officers and the RVCs, depicted in Figure A1, begins with a move of Nature, selecting the former security agent's type. With probability cm $(0 < cm < 1)$ the type is desirable and with probability $1 - cm$ the type is undesirable[19]. Upon learning his type (but not knowing the RVC's cost of vetting) the agent decides whether to apply for verification. If the agent refrains from applying, he receives the outside option of o, regardless of his type, and the

[19] Note that since we learned from the decision-making model that the effects of c and m are interchangeable, we will focus on the joint movement of c and m in the desirable and undesirable direction

Payoffs to RVC	positively verifying competent moderate	1
	positively verifying incomp. or extreme	1-d
	firing agent	1-t
	cost of carrying out selective purge	s

Payoffs to Officer	competent & moderate	incomp. or extreme
approved by RVC	b	b
denied by RVC	o	γo
not applied	o	o

Table A1: Payoffs at terminal nodes of the game

game ends. If the agent decides to apply for verification, their application is made visible to the RVC, although the RVC does not observe the type of the agent. Upon observing the application, the RVC can:

1. fire the agent without reading the personnel file or interviewing him (scaled up, this decision corresponds to a thorough purge),
2. rehire the agent without reading the file (scaled up, this decision corresponds to "no purge"),
3. at a cost (s) carefully read the file and learn the type of the applicant (this action, regardless of the outcome, scaled up corresponds to a selective purge).

In the event of a firing, the desirable type receives the same outside option, o, as he would if he never applied. The undesirable type, however, receives the outside option decreased by a factor γ, $0 < \gamma < 1$. This is to account for the fact that valuable job searching time was wasted while incompetent or extreme types underwent verification and the pool of lucrative jobs for these types has shrunk. The justification for this assumption is that after the lapse of time it takes for the verification process to play out, employers can afford to be more selective about who they chose to hire. A second way through which an undesirable officer can lose his job is through vetting, but this action is costly for the RVC (the RVC pays a private cost s for embarking on a selective purge) and with probability $1 - q$, despite paying the cost, vetting may still fail to uncover an undesirable agent. When an officer is re-hired, he gets the payoff of b, $b > o$, the benefit of being re-hired in security services is higher than the outside option. The desirable officer will receive that payoff in the event of a selective purge as well (we exclude the possibility of false evidence of incompetence or extremeness). The RVC gets a payoff of 1 from rehiring a desirable type, but a payoff of $1 - d$ if it rehires an undesirable type (note, that in the decision model, the payoff from hiring this type was just zero, that is d was set at 1). If it fires the agent and has to train a new one, its payoff is $1 - t$ and we assume $t < d$, that is the cost of training is lower than the cost of hiring an undesirable agent. The game is formally described in Appendix **??** and Table A1 summarizes the payoffs.

Even though the term "purge" implies a decision regarding the entire pool of applicants, not an individual applicant, the only information at the RVC's disposal when evaluating a specific officer is whether he applied to be screened for reemployment, its cost of vetting, and parameters of the game that can only be assessed at the general wojewodztwo level. Therefore, we will use the terms thorough purge and "fire" on the one hand and "rehire" and "no purge" on the other interchangeably.

This is a signaling game with costly information (Spence 1978; Banks 2013). The signal here

is the decision to apply for verification. Applying is more costly for the undesirable type because he forgoes a potentially more lucrative employment opportunity.

We solve the model, as is customary for costly signaling games, for Perfect Bayesian Equilibrium (PBE). The key feature of PBE is that in addition to strategies in equilibrium, one must also specify beliefs that support that equilibrium. We follow the PBE definition from (Gibbons 1992) and assume that (1): at each information set players must have a belief about which node they have reached in the information set (2): the action taken at each information set must be optimal given players' belief at that information set and their subsequent strategies (3): at all information sets on the equilibrium path, beliefs are determined by Bayes rule and players' equilibrium strategies and (4): beliefs are also rational (according to Bayes) off the equilibrium path.

As explained in the main text, equilibria in signaling games can be roughly divided into separating and pooling. In the former, the player who moves last (the RVC) updates their beliefs relative to the priors by learning something about the type of the player who moves first (the agent). This learning need not be complete, in which case we are dealing with semi-separating equilibria. There is also a class of pooling equilibria, where no learning takes place: the beliefs of the uninformed player are the same as they were in the beginning of the game.[20]

We first show that this game has no pure separating equilibria and explain what this means for the conventional wisdom that motivates the creation of vetting commissions. Next, we summarize our finding regarding pooling and semi-separating equilibria and explain what this implies for the empirical predictions from this more general model.

A.1.1 The existence of pure separating equilibrium

The conventional wisdom about the operation of verification commissions implicitly posits the existence of a separating equilibrium in which all the desirable types apply, but undesirable ones do not, while the RVC conducts a selective purge. In the main text we explained why such a pure separating equilibrium equilibrium does not exist. We can also formally show why the second separating equilibrium—where only undesirable types apply, while the desirable ones do not—does not exist.

If such an equilibrium were to exist, we would have to assume that when the RVC sees fewer applicants than eligible candidates it believes such applications come from the undesirable types. But the payoff associated with hiring them is $1 - d$ while the payoff associated with vetting them is $(1 - d - s)(1 - q) + (1 - t - s)q$, both of which are lower than the payoff from firing them: $1 - t$, in light of the fact that $d > t$. But firing them gives undesirable types who apply γo, as opposed to o they would have gotten have they not applied in the first place. These two arguments lead to the proposition:

Proposition 1. *No pure separating equilibria exist in the Verification Game*

Following this reasoning we are left with just two kinds of equilibria: semi-separating and pooling. The comparison of the pooling equilibria with the semi-pooling are the most interesting from the point of view of the robustness of the decision model discussed in section 3.1. If pooling equilibria exist for wider parameter spaces than semi-separating equilibria, we can comfortably ignore the change in opportunities for the desirable (o) and undesirable (γo) agents created by applying for verification. In a nutshell, we can *assume that both types of agents apply at the same*

[20] Note that in our model, although the RVC also has private information (about the cost of vetting, s), the agent has no opportunity to update his beliefs about s because the RVC moves second. Hence, the verification game is a model with just one-sided updating.

rate and if they apply at the same rate, the signaling model is redundant. Moreover, pooling equilibria are relevant for our purposes because they can be tested with aggregate data (as the individual agent decision is always to apply, regardless of their type).

A.2 Pooling equilibria

In pooling equilibria both desirable and undesirable types apply. Because of this, the RVC does not update its prior beliefs about the type it faces. The beliefs supporting this equilibrium are given by:

$$Pr(desirable|apply) = cm$$

$$Pr(undesirable|apply) = 1 - cm$$

Before characterizing the pooling equilibrium, we note that a strategy of the RVC is a function of s, the privately observed costs of verification. To characterize the best response strategy, we begin by asking when does the RVC vet when both types apply. In such a situation, given the beliefs above, it must be the case that the RVC prefers vetting to both firing and hiring:

$$EU_{RVC}(vet; apply, apply) \geq EU_{RVC}(fire; apply, apply) \tag{4}$$

and

$$EU_{RVC}(vet; apply, apply) \geq EU_{RVC}(hire; apply, apply) \tag{5}$$

Substituting $EU_{RVC}(vet; apply, apply) = (1 - cm)[q(1 - t - s) + (1 - q)(1 - d - s)] + cm(1 - s)$ and $EU_{RVC}(fire; apply, apply) = 1 - t$ into equation 4, we obtain:

$$s \leq s^F \equiv cmd - (d - t)(1 - q(1 - cm)) \tag{6}$$

Substituting $EU_{RVC}(fire; apply, apply) = 1 - t$ and $EU_{RVC}(hire; apply, apply) = (1 - cm)(1 - d) + cm$ into equation 5, we obtain:

$$s \leq s^H \equiv (d - t)q(1 - cm) \tag{7}$$

Whether $s^H > s^F$ or $s^F > s^H$ or $s^F = s^H$ (implying whether condition 6) or condition (7) is binding) will depend on the specific values of cm, d, and t.

In Figure 5 of the main text, the constraints s^H and s^F correspond to the lines (in green and red, respectively). Case 1 corresponds to the area where the green line is below the red (the condition on s^H is binding), and Case 2 to the area when the red line is below the green (the condition on s^F is binding). The fact that one line slopes downwards, while the other slopes upwards has interesting consequences for when vetting is optimal.

When the cost of verification s does not satisfy either 7 or 6, to find the RVC's best response to the officer's $(apply, apply)$, we must also compare $EU_{RVC}(fire; apply, apply)$ with $EU_{RVC}(hire; apply, apply)$. We see that

$$EU_{RVC}(hire; apply, apply) > EU_{RVC}(fire; apply, apply) \text{ iff } t > d(1 - cm) \tag{8}$$

However, it is easy to see that (8) implies $s^F > s^H$, while failure to satisfy (8) implies $s^H > s^F$. This is also plainly visible in Figure 5, where the point at which the red line (representing s^F) and the green line (representing s^H) cross coincides with either "Hire" or "Fire" being optimal for the RVC in the event that the costs of vetting are too high.

This reduces the number of cases we have to consider to only two. Figure A2 summarizes the theoretically possible cases, showing that only 1a and 2b survive:

1a $t > d(1 - cm)$

2b $t < d(1 - cm)$

In case 1a, the best response of the RVC is a strategy s^*, which (in case 1a (see Figure A2) takes the form:

$$s = \begin{cases} \text{vet} & \text{if } s \leq s^H \\ \text{hire} & \text{if } s \geq s^H \end{cases} \tag{9}$$

For a pooling equilibrium with s^* and applying to hold, we need the expected utility of both types of officers from applying to exceed the expected utility of not applying. For the desirable type, this is trivially satisfied, as $b > o$ and the same type of employment opportunities await him before and after the verification process. However, for the undesirable type, to make sure that applying is incentive compatible, we must compare his expected utility from not applying, o, to his expected utility from applying, the latter of which is contingent on RVC's best response function (since the agent does not knoow s):

$$EU_{\,cm}(apply|s^*) > EU_{\,cm}(\sim apply|s^*) \tag{10}$$

which in case **1a**, is equivalent to:

$$Pr(s < s^H)[(1 - q)b + q * \gamma o)] + Pr(s \geq s^H)b \geq o$$

This is turn becomes

$$s^H((1 - q)b + q\gamma o) + (1 - s^H)b \geq o \tag{11}$$

In terms of constraints on γ, this can be written as expression[21]: $\gamma = \bar{\gamma} \geq \frac{o - b(1 - s^H q)}{s^H q o}$.

To see that this is less than one, as assumed, note that the denominator is larger than the numerator because $o > b$. Similar inequalities have been designed and solved for case 2b and are reported in Appendix A1 and are summarized in the Figure A2. The condition on γ that has to be satisfied in Case 2b is $\gamma \geq \bar{\gamma} = \frac{o - s^F b(1-q)}{o(1 - s^F(1-q))}$.

We summarize the pooling equilibrium, in our next proposition.

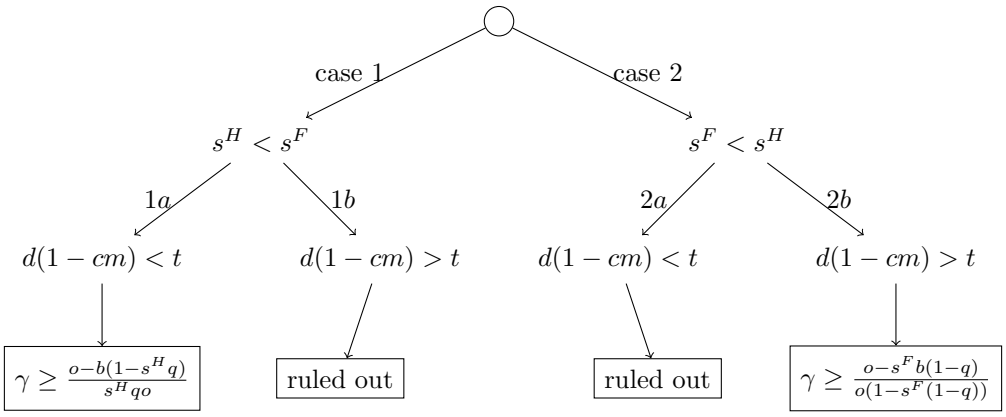
Proposition 2. *Assume that opportunity costs of the undesirable type applying for verification are sufficiently low ($\gamma \geq \bar{\gamma}$). Then, there exists a pooling equilibrium where both types apply, while the RVC vets if the cost of verification is sufficiently low ($s < \min\{s^F, s^H\}$). If the cost of verification is not low, the RVC fires when the proportion of desirable agents is sufficiently low ($cm < \frac{d-t}{d}$), and hires if the cost of vetting proportion of desirable agents is sufficiently high ($cm > \frac{d-t}{d}$). Vetting is easiest to achieve for middling shares of desirable agents.*

The remainder of this appendix is made up of three subsections. In the first, we solve for the incentive compatibility constraints of the undesirable agents in the above case 1a & 2b, that is case that was not solved for in the main text. The following subsection considers equilibria in mixed strategies, where it is the undesirable agent who is doing the mixing. The final subsection presents several additional comparative statics.

[21] Reported also in Figure A2

A.3 Incentive compatibility of undesirable agents in case 2b

Figure A2: Incentive compatibility in Pooling equilibrium



Recall from our discussion in section 4.2. that incentive compatibility in a pooling equilibrium requires that, in addition the desirable type's preferring applying for verification to not applying (which we argued is trivially satisfied, the undesirable type must prefer it too, that is:

$$EU_{\ cm}(apply|s^*) > EU_{\ cm}(\ apply|s^*) \tag{12}$$

In case **1a**, this is equivalent to:

$$Pr(s < s^H)[(1-q)b + q * \gamma o)] + Pr(s \geq s^H)b \geq o$$

$$s^H((1-q)b + q\gamma o) + (1-s^H)b \geq o \tag{13}$$

.

Solving for γ produces the expression reported in Figure A2: $\gamma \geq \frac{o-b(1-s^H q)}{s^H qo}$. This is the incentive compatibility constraint for 1a.

Since cases 1b and 2a are eliminated, we next consider case **2b**, where the incentive compatibility condition described in 12 becomes

$$Pr(s < s^F)[(1-q)b + q * \gamma o)] + Pr(s \geq s^F)\gamma o \geq o$$

, which in light of s's uniform distribution reduces to:

$$s^F[(1-q)b + q * \gamma o] + (1-s^F)\gamma o \geq o$$

Once more, solving for γ yields $\gamma \geq \frac{o-s^F b(1-q)}{o(1-s^F(1-q))}$, as reported in Figure A2. This constitutes the incentive compatibility constraint for case 2b.

A.4 Semi-separating equilibrium

The previous section solved for incentive compatibility, interpreted as the maximal opportunity costs associated with applying for verification by incompetent or extreme types. The question that

arises is what kind of equilibrium obtains when these opportunity costs are too high. We conjecture that the outcome would be a pair of strategies that defines a mixed strategy equilibrium with the mixing player being the undesirable type. This is intuitive, as the desirable type has a dominant strategy, while the RVC's strategy seen from the point of view of the agent is itself a probability distribution over the actions of "vet", "fire", and "hire."

To find this equilibrium, we first calculate what the updated beliefs of the RVC would be that they are facing desirable versus undesirable agent.

Assuming that the undesirable type applies with probability λ and refrains from applying with probability $1-\lambda$, these beliefs are best expressed as:

$$Pr(desirable|apply) = \frac{cm}{cm + \lambda(1-cm)}$$

$$Pr(undesirable|apply) = \frac{\lambda(1-cm)}{cm + \lambda(1-cm)}$$

Next we calculate the RVC's expected utility given these beliefs for "vet", "fire", and "hire":

$$EU_{RVC}(vet|apply, \lambda) = Pr(desirable|apply)*1+Pr(undesirable|apply)*[q(1-t)+(1-q)(1-d)]-s \tag{14}$$

$$EU_{RVC}(hire|apply, \lambda) = Pr(desirable|apply) * 1 + Pr(undesirable|apply) * (1-d) \tag{15}$$

$$EU_{RVC}(fire|apply, \lambda) = 1 - t \tag{16}$$

After substituting the updated beliefs into 14, 15, and 16, we set 14 equal to 15 to find the cutoff $s^{H(mix)}$ and we set 14 equal to 16 to find the cutoff $s^{F(mix)}$.[22]

This way we obtain:

$$s^{H(mix)} = \frac{\lambda(1-cm)q(d-t)}{cm + \lambda(1-cm)} \tag{17}$$

and

$$s^{F(mix)} = \frac{-1 + t + cm + \lambda(1-cm)(1-qt-d(1-q))}{cm + \lambda(1-cm)} \tag{18}$$

This ensures that at the cutoffs, the RVC is indifferent between vetting and firing on the one hand and between vetting and hiring on the other hand.

In the final step, we need to ensure that given these cutoffs, the undesirable type is indifferent between refraining to apply and applying, that is:

$$EU_{\sim CM}(\sim apply) = EU_{\ CM}(apply|s^{mix}) \tag{19}$$

The expected utility of not applying is always o, but the expected utility of applying depends on the ordering of the cutoffs and on whether when vetting is not optimal, the RVC prefers firing to hiring.

[22] This approach follows Fudenberg and Tirole (2020, p.213-19) technique for finding equilibria in mixed strategies in games with incomplete information and at least one continuous type.

In order to find the latter, we need to calculate the condition $EU_{RVC}(fire|\lambda) \geq EU_{RVC}(hire|\lambda)$. Substituting 16 and 15 into this inequality, we arrive at

$$t \leq \frac{d\lambda(1-cm)}{cm+\lambda(1-cm)} \tag{20}$$

In this mixed strategy equilibrium, as in the pure strategy equilibrium (see Figure 5), the line defining $s^{F(mix)}$ is an increasing function of cm, while the line defining $s^{H(mix)}$ is a decreasing function of cm. Below both lines, RVC vets; above the $s^{F(mix)}$ line (but below the $s^{H(mix)}$ line) it fires; and above the $s^{H(mix)}$ line (but below the $s^{F(mix)}$ line) it hires. What does it do above both lines depends on whether condition 20 is satisfied or not.

In light of this, four cases need to be considered. They are defined by whether condition 20 is satisfied and whether $s^{F(mix)} > s^{H(mix)}$ or the other way around. A graph representing these cases would look just like Figure A2, except that only one of the cases is ruled out (condition 20 is only compatible with $s^{H(mix)} > s^{F(mix)}$, as $t \leq \frac{d\lambda(1-cm)}{cm+\lambda(1-cm)}$ implies $s^{H(mix)} > s^{F(mix)}$)

A.4.1 Low cm

We begin by considering $s^{F(mix)} < s^{H(mix)}$ and $t \leq \frac{d\lambda(1-cm)}{cm+\lambda(1-cm)}$ In this case, equation 19 can be written as

$$o = Pr(s < s^{F(mix)})EU_{\ CM}(vet, apply) + Pr(s > s^{F(mix)})EU_{\ CM}(fire, apply)$$

Since s is still uniformly distributed, this reduces to

$$o(1-\gamma) = s^{F(mix)}(1 - q(b - \gamma o)$$

Finally, we substitute the expression 18 into the above to obtain:

$$o(1-\gamma)(cm + \lambda(1-cm)) = (1-q)(b - \gamma o(-1 + t + cm + \lambda(1-cm)(1 - qt - d(1-q)))$$

Solving the above expression for λ provides the value of the exact probability with which undesirable types should apply for verification in this semi-separating equilibrium. The general solution of this equation, however, requires recursive methods, so it is easier to solve after making certain assumption about our parameters. If we were to use the parameters used to create Figure 5 and assuming additionally that $\gamma = .75, o = .15, q = .75$ and $b = 2$, $\lambda < 0$, which is impossible given that λ must be a probability.

A.4.2 Remaining cases

The second case we consider is $s^{F(mix)} < s^{H(mix)}$ and $t \geq \frac{d\lambda(1-cm)}{cm+\lambda(1-cm)}$. In this case, equation 19 can be written as

$$o = Pr(s < s^{F(mix)})EU_{\ CM}(vet) + Pr(s^{F(mix)} > s > s^{H(mix)})EU_{\ CM}(fire) + Pr(s > s^{H(mix)})EU(hire)$$

Since s is still uniformly distributed, this reduces to

$$b - 0 = s^{H(mix)} - s^{F(mix)}(1-q)(b - \gamma o)$$

Again, after substituting for $s^{H(mix)}$ and $s^{F(mix)}$ using equations 18 and 17, we can solve for λ. In this instance, λ is indeed between 0 and 1, thus a semin-separating equilibrium exists.

The third case to consider is $s^{F(mix)} \geq s^{H(mix)}$ and $t \geq \frac{d\lambda(1-cm)}{cm+\lambda(1-cm)}$. In this case, equation 19 can be written as

$$o = Pr(s < s^{H(mix)})EU_{\ CM}(vet) + Pr(s > s^{H(mix)})EU_{\ CM}(hire).$$

Substituting into the above equations for $s^{H(mix)}$ using equation 17 and substituting in for the expected utilities 14 and 15: yields

$$\lambda = \frac{cm(b-o}{(1-cm)q^2(d-t)(b-\gamma o) - b + o}$$

for the third case.

B Additional empirical evidence

This section provides additional qualitative evidence for the claims we make in the main text, as well as introduces alternative operationalization of the key variables.

B.1 Qualitative evidence for claims in the main text

Our formal models made several assumptions that based on our archival research, we were confident in making. However, in the main text, due to space limitation, we could only offer select evidence supporting these assumptions. This appendix offers additional evidence for our claims.

We begin with evidence of higher competence of secret police in Poland relative to uniformed police: Oseka (2008) analyzing applications to the communist enforcement agencies, established that although applicants for entry-level positions in the militia came from rural areas and rarely possessed even a high school diploma, this was no longer the case for applicants to the secret police. The latter were selected from among the best and brightest of the militia, which had often educated them through high school, and sometimes even college, level. An internal survey accompanying exit interviews (Komar and Niedzialek 1990) indicates that work in the Polish secret police generally turned uneducated, poor and rural unskilled workers into educated professionals, living in their own urban apartments. Hence, work for the secret police was appealing to the relevant recruitment pool of militia officers. Typically, the best militia officers, rather than the worst, would be transferred from the regular militia forces to the secret police forces. To merit such transfer, they had to exhibit potential for learning the most valuable skill a secret police officer could acquire: the ability to recruit informers. This was a complex task and one on which the secret police agencies spent considerable resources, particularly in areas where the battle for citizens' hearts and minds was most heated (Piotrowska 2020).

B.2 File destruction

The next set of archival sources we describe offers evidence of file destruction and the randomness of file destruction (recall, that in our signaling model, this was represented by parameter q) Because the Polish transition to democracy was gradual, the security services and their collaborators had the time to prepare for its consequences and acts of transitional justice. One crucial act was the destruction of evidence of secret police operation. What is key for our argument is (i) whether the work files of officers were destroyed, (ii) to what extent were they a priority and (iii) whether there is anything we can say about the extent to which the officers knew about their file survival.

While the eradication of files of the unofficial informants has received the most attention and was the highest priority, officers also had incentives to destroy any information that incriminated the secret police itself. This became more urgent with the formation of the Extraordinary Sejm Commission to Investigate the Activities of the Ministry of the Interior on August 17, 1989. The objective of the commission was to investigate 93 cases of deaths presented by the Helsinki Committee in Poland, which the security organs of the Polish People's Republic were suspected of. Hence, the operational files related to these cases are very likely to be incomplete. However, it is not unreasonable to assume that the files of the key officers involved could have been equally purged of incriminating information.

However, while we have a motive for officer file destruction, we do not have direct evidence for this, nor can we assume that the individual officers knew whether their files survived the transition. This is because the bulk of the activity was placed on the high-profile cases and it is unclear if individuals could curate their own files.

B.3 Example of a form

An example of a page from the files constituting our data is provided in Figure B3.[23]

B.4 Alternative operationalization of the variables

In this section we consider alternative operationalizations of the key variables analyzed in the comparison of means.

B.4.1 Classification robustness checks

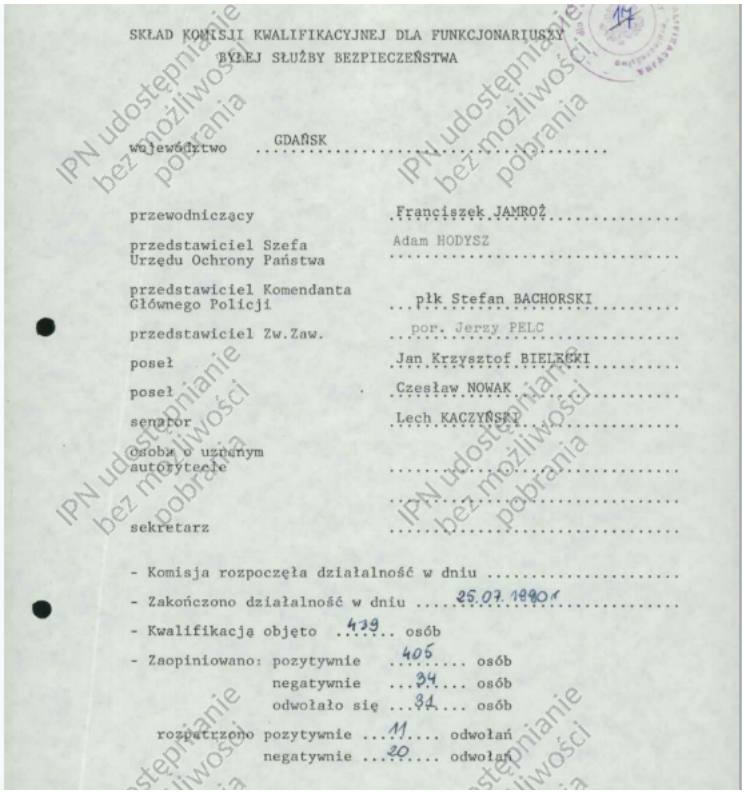
Letters and reports from the archives of RVCs depicting commissions' work corroborate our classification. In wojewodztwa classified as having "no purges", the documents consist of complaints regarding their low severity. For instance, Wloclawek's Solidarity Trade Union rep wrote a letter to the RVC complaining about the high level of positive verifications in the wojewodztwo (BU/3546/52). In Zamosc, the commission wrote "Society judged our position as too lenient. SB's activity in the Zamosc area was particularly brutal and repressive." (BU/3546/53 p.22). Finally, in Radom, the leader of the RVC wrote "We feel that the Commission's activities have not met public expectations." (BU/3546/39).

Documents from wojewodztwa that saw selective purges show fewer complaints. As described in the paired comparison section, in wojewodztwa that saw selective purges, such as Bydgoskie and Tarnowskie, there is evidence that the RVCs could take their time considering each case thoroughly (IPN By/453/47; IPN BU/3546/48)

Correspondence stored in the archives of the wojewodztwa that underwent thorough purges strikes a yet different tone, suggesting a desire for revenge. According to the author (a former SB

[23] The form shows the composition for each district (wojewodztwo), including the head (przewodniczacy), an UOP representative (przedstawiciel Szefa Urzadu Ochrony Panstwa), a local police representative (przestawiciel Komendanta Glownego Policji), a police union rep (przedstawiciel zw. zaw.), MP (posel), senator, individual(s) with high authority (osoba o uznanym autorytecie), secretary (sekretarz). It also gives the dates of beginning (Komisja rozpoczela dzialalnosc w dniu) and end (Zakonczono dzialalnosc w dniu) of its operation and the verification outcome: the number verified (Kwalifikacja objeto), verified positively (zaopiniowano pozytywnie), negatively (negatywnie), and those who appealed the outcome (odwolalo sie).We investigate the consequences of such appeals in separate work. Here, it suffices to say that the possibility of appeal was not anticipated by RVCs.

Figure B3: Example of a Regional Verification Commission form summarizing the commission's decisions



SKŁAD KOMISJI KWALIFIKACYJNEJ DLA FUNKCJONARIUSZY
BYŁEJ SŁUŻBY BEZPIECZEŃSTWA

województwo GDAŃSK

przewodniczący Franciszek JAMROŻ

przedstawiciel Szefa Urzędu Ochrony Państwa Adam HODYSZ

przedstawiciel Komendanta Głównego Policji płk Stefan BACHORSKI

przedstawiciel Zw.Zaw. por. Jerzy PELC

poseł Jan Krzysztof BIELECKI

poseł Czesław NOWAK

senator Lech KACZYŃSKI

osoba o uznanym autorytecie

sekretarz

- Komisja rozpoczęła działalność w dniu .....
- Zakończono działalność w dniu 25.07.1990 r.
- Kwalifikacją objęto 439 osób
- Zaopiniowano: pozytywnie 405 osób
  negatywnie 34 osób
  odwołało się 31 osób
  rozpatrzono pozytywnie 11 odwołań
  negatywnie 20 odwołań

officer) of a letter addressed to the Ombudsman for Human Rights "on its first day of proceedings, the commission [RVC in Tarnobrzeg] handled a vast majority of valuable and highly educated personnel. (...) Moreover (...) there was a list of 40 names drawn up by "Solidarity" demanding they be dismissed. Among those negatively verified were secretaries, an archivist, and even an officer who served on one of the new parliamentary committees (...) This was but another act of revenge carried out in our country" (BU/3546/47 p.41).

As a final robustness check of our categorization, we use information about officers that appealed their RVC decision along with the results of appeals. Our purge classification would be supported if the proportion of appeals were highest in wojewodztwa that we classified as implementing thorough purges and lowest in wojewodztwa with "no purges". Moreover, thorough purge wojewodztwa should have the most decisions reversed on appeal.

A comparison of means shows exactly that. In wojewodztwa with thorough purges 91.9% officers verified negatively had appealed the result, while in wojewodztwa with selective purges the corresponding number was 90.4%. Finally, wojewodztwa with "no purges" had only 85.4% of the rejected officers appeal their decisions. Of these appeals, the highest proportion— 75.5%— of reversals took place in wojewodztwa experiencing thorough purges, followed by 60.7% in wojewodztwa with selective purges. The corresponding figure for wojewodztwa with "no purges" was 58.7%. These additional robustness checks bolster the credibility of our classification rule.

B.4.2 Median based classification

First, we use the same model variables, but consider the means when the purge categories are defined using median, rather than mean, variables.

To do so, recall that the main categorization considers a wojewodztwo as having a selective purge if the time devoted to each application was higher than the mean (Table 1). A thorough purge is defined as a situation when the time was short (lower than the mean) and the proportion of those verified negatively high (higher than the mean), and "no purge" is associated with a similarly short time, but a low proportion of those verified negatively. In what follows, we follow a similar rule, but using the median, rather than mean, values of the variables.

Table B2: Median-based classification of purges

Purge type	Proportion verified negatively	Hours per application	Frequency	Percent
No	<0.51	<11.3	10	23.81
Selective		>11.3	21	50
Thorough	>0.51	<11.3	11	26.19

Because the median time is significantly lower than the mean, using median values classifies more wojewodztwa as having experienced a selective purge.

B.4.3 Alternative specifications of explanatory variables

Beside the main classification, all the theoretically key variables can be operationalized in alternative ways.

Starting with moderation (m) (1-repression) and the cost of running an RVC (s), recall from the main text that repression is measured as:

1. Number of those sentenced during the Martial Law/ Solidarity membership
2. Number of those tried during the Martial Law/ Solidarity membership

Table in the text uses the first operationalization. Similarly, the cost of setting up the commission can be measured in three ways: 1.Population density 2.Average salary in a given wojewodztwo 3.Revenue per capita of a given wojewodztwo (excluding transfers from the central government).

To further explore the robustness of the comparison of means, we consider the means of the various operationalizations in wojewodztwa classified as having experienced different purge types (here we use the mean values). Table B3 shows that the relative relationship between the means remains the same regardless of the specification. Most importantly, in all three operationalizations, the cost of setting up the RVC is the lowest in wojewodztwa classified as having seen a selective purge.

Moreover, we propose a third alternative operationalization of extremeness that does not rely on cases brought before court during the Martial Law: the number of disappearances and murders attributed to the SB in the 1980s, aggregated on a wojewodztwo level. The average number of victims per wojewodztwo was. 1.96 with a range from 0 to 14 (Table 2). Tarnowskie had no reported victims, while Rzeszwoskie had 1. Similarly, Wloclawskie had no victims, but Bydgoskie had 4.

Table B3: Means of alternatively operationalized key explanatory variables by purge type

	m_1	m_2	s_1	s_2	s_3
no purge	.999858	.999772	.198	198.37	259.86
selective purge	.999851	.999739	.089	192.58	229.66
thorough purge	.999858	.999772	.164	200.98	250.09

Table B4: Alternatively operationalized key explanatory variables for paired comparison

	r_1	r_2	s_1	s_2	s_3
Tarnowskie	0.0000705	0.0001795	0.160443	188.1	222.1817
Rzeszowskie	0.0002542	0.0003583	0.1631	183.5	245.67
Bydgoskie	0.0002925	0.0005408	0.1069	198.9	289.6
Wloclawek	0.0000778	0.0001778	0.09743	191.8	200.2238

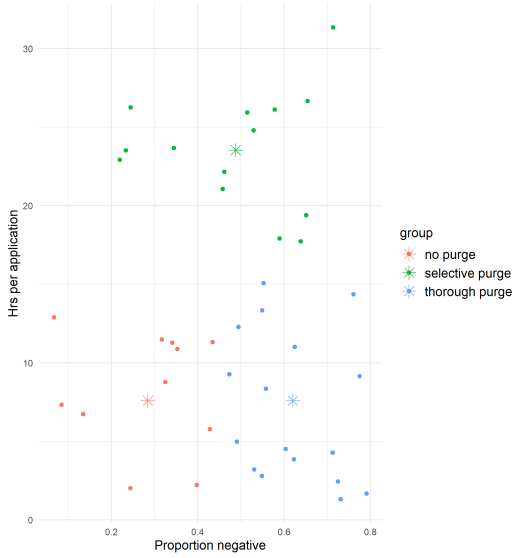
Finally, moving on to measuring *competence*, in addition to the main measure (the proportion of officers hired by the department for the Protection of the Economy), we consider an alternative: the proportion of them who, during their career, took additional courses and workshops. We source this data from IPN (https://katalog.bip.ipn.gov.pl/funkcjonariusze/?catalog=5), considering those working for SB in 1990, and aggregating the proportion who took additional courses on a wojewodztwo level. On average 0.42 officers took additional courses, with a range from 12% to 75%. Tarnowskie (0.45) and Rzeszowskie (0.31) had both relative low rates of additional schooling. Bydgoskie (0.68) and Wloclawskie (0.55) saw relatively high levels of training.

B.5 K-means algorithm

One alternative approach to classify the wojewodztwa is to use the K-means algorithm, a unsupervised machine learning approach used to subset the dataset into K exhaustive and exclusive clusters. In this process, every data point is allocated to the cluster whose mean or centroid is closest to it. Its key advantage for us is that it does not rely on our judgment in dividing the sample into the three clusters (three types of purges).

Figure B4 shows the results of the classification using the K means algorithm, including the centroids marked with a star.

Figure B4: Classification of purge types using K-means clustering



Note The centroids are marked with a star.

Regardless of the exact specification, the classification of the majority of wojewodztwa (and importantly those we use in our paired comparison) is not sensitive to the exact cutoffs we adopt to define the different purge types.

B.6 Pairs of wojewodztwa

The distance measure we used is the Euclidean distance, which is a measure of the bird's-eye distance between two points in Euclidean space.

For instance, a pair of points with coordinates (x_1, y_1, z_1) and (x_2, y_2, z_2), the Euclidean distance δ between them is calculated as:

$$\delta = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \tag{21}$$

In our case, the coordinates x, y, and z correspond to the values of the variables population density, moderation, and competence for a given wojewodztwo.

This process creates a list of pairs of wojewodztwa, with the five closest being:

1. **Rzeszowskie and Tarnowskie** (Distance: 0.003390265)
2. **Bydgoskie and Wloclawskie** (Distance: 0.013007807)
3. Ciechanowskie and Olsztynskie (Distance: 0.013119470)
4. Leszczynskie and Zielonogorskie (Distance:0.018498317)
5. Bialostockie and Lomzynskie (Distance: 0.022428315)

To see that the wojewodztwa are very similar also according to alternative specifications of the explanatory variables, consider Table B5. The alternative specifications tell a similar story: the two pairs are very similar in their population density and competence levels (defined in two ways) and they saw varying levels of extremeness. Moreover, the competence of Tarnowskie and Rzeszowskie was low relative to the other pair, regardless of the exact measure.

Table B5: wojewodztwa compared and their characteristics

Wojewodztwo	Repression (court)	Repression (murders)	Population density	Competence (PoE)	Competence (courses)
Tarnowskie	0.0000705	0	0.160	0.211	0.451
Rzeszowskie	0.0002542	1	0.163	0.213	0.305
Bydgoskie	0.0002925	4	0.107	0.363	0.679
Wloclawskie	0.0000778	0	0.097	0.354	0.546

C Out-of-sample test: Ukraine

The subnational analysis presented in the paper gives us high confidence in the internal validity of our results and their potential for explaining the dynamics of security service reform in Poland in the 1990s. But what insights does it provide beyond that time and place? How representative of screening mechanisms in the former Soviet Bloc are Poland's verification commissions?

We explore the potential shadow case of Ukraine, with whom Poland shares a post-communist context. When in 2014, the world turned its attention to Euromaidan in Kyiv, Ukraine, it was in response to the brutal pacification of peaceful protesters in the central square of Kyiv, the Maidan. People had gathered there in opposition to the ruling government's pulling out of a cooperation agreement with the EU in what was perceived as a pro-Russia move. The government led by Viktor Yanukovych was in the last months of what was generally perceived as his final term. After days of trying to deflect the protest, Yanukovych resorted to sending the most brutal troops of riot police, the infamous Berkut unit, to pacify the demonstrators. Police fired on the occupants of Maidan and as a result of police actions, more than a hundred protesters died and many more suffered injuries. Europe watched in shock as a peaceful protest turned bloody. The Revolution of Dignity that followed events of the Euromaidan eventually ousted Yanukovych's government and led to the creation of a new interim government followed by new presidential elections that put in power Petro Poroshenko. Police reform was at the very top of Poroshenko's agenda. Although following the break-up of the Soviet Union, Ukraine had not purged its security agencies of agents of repression, the interim government disbanded the violent 14,000 troops-strong Berkut unit. Several members of the unit fled to by now Russian-annexed Crimea asking for asylum in exchange for allegiance (which likely explains their eagerness to violently suppress Euromaidan). In 2022, during Russian War in Ukraine, troops from Berkut units were seen fighting on the side of Russia.

But the disbanding of Berkut was only one part of the security agencies reform. What followed under the new government was a reform comprising of (1) the creation the newly founded (on principles of democratic policing) National Police of Ukraine; (2) the establishment of a Patrol

Police staffed by newly trained recruits; (3) and a verification process resembling the Polish process (but called "re-attestation" or "re-certification") of almost 70,000 police officers.

Ukrainian re-attestation commissions included trusted civil society members. Their decisions were based on studying officers' personnel files and interviews with officers, sometimes with the use of a polygraph. Initially, 7.7% of the police force was fired, including 27% officers in leadership positions. However, the process was rumored to have succumbed to capture by insiders of the previous regime who added an appeals stage to the process, as a result of which 93% of the initially fired were rehired.

In the terminology of our model, the police reform in post-Euromaidan Ukraine is a combination of a selective purge (in the initial phases) and "no purge" (at the appeals stage). Consequently, the process in Ukraine lends itself to analysis with the tools of our model of police reform. First, note that disbanding of the Berkut unit (a thorough purge) would have been a good choice had the Berkut forces been both extreme and incompetent. While the extremism of Berkut troops is evident from its troops' readiness to declare allegiance to Russia following the annexation of Crimea and during the ongoing war, incompetence is harder to establish in light of the "elite" status of Berkut. In the late Soviet period, Berkut forces were among the best-trained units in the former Soviet region (Marat 2018). Hence, these troops had skills that would not be repurposed in the eventuality of a thorough purge. Perhaps because of such considerations, after the unit was disbanded, nearly 41% (1,641) of Berkut officers joined the new Ukrainian police force and consequently, the process resembled a selective, rather than thorough purge. While, as remarked above, considerable numbers of the former Berkut pledged allegiance to Russia, some pro-Ukrainian Berkut officers also joined the National Guard of Ukraine and ended up fighting alongside the regular Ukrainian Army against pro-Russian separatists in Donbas. Second, retaining the leadership of the security forces and the police helped preserve the culture of these institutions from the Soviet era. Hence, despite purges of extreme and incompetent actors lower in police hierarchy and the training of new recruits, officers of the old regime preserved the corrupt culture. In sum, the appeals process nullified the impetus of the initial reform.

Supplemental appendix material is available at the link:
https://doi.org/10.17605/OSF.IO/6HYUG