

Gathmann, Christina; Boyer, Pierre

**Conference Paper**

## Trial and Error? Reelection Concerns and Policy Experimentation during the US Welfare Reform

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung - Theorie und Politik - Session: Public Choice and Welfare, No. G12-V3

**Provided in Cooperation with:**

Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Gathmann, Christina; Boyer, Pierre (2015) : Trial and Error? Reelection Concerns and Policy Experimentation during the US Welfare Reform, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung - Theorie und Politik - Session: Public Choice and Welfare, No. G12-V3, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft

This Version is available at:

<https://hdl.handle.net/10419/112844>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Trial and Error? Reelection Concerns and Policy Experimentation during the US Welfare Reform\*

Andreas Bernecker<sup>†</sup> Pierre C. Boyer<sup>‡</sup> Christina Gathmann<sup>§</sup>

This version: November 7, 2014

## Abstract

We study the political determinants of policy experimentation during the US welfare reform. Among other changes, this reform shifted the autonomy to implement welfare policies from the federal government to the individual states. In line the predictions from a political agency model, we find that (i) governors with high initial reputation among voters experiment less with welfare policies; (ii) governors with lower reelection concerns experiment more; (iii) governors with little experience are more likely to revert an experiment, but are also more likely to stick to a policy experiment with high potential gains. Overall our findings suggest that reelection concerns play an important role for policy experimentation and reversals.

Keywords: Policy innovation; reputation concerns; US welfare reform.

JEL classification codes: I38; H11; H77; D78; D83.

---

\*We thank Toke Aidt, Manuel Arellano, Henning Bohn, Matz Dahlberg, Peter Fredriksson, Vincenzo Galasso, Mario Jametti, Dylan Minor, Espen Moen, Jörg Oechssler, Paanu Poutvara, Torsten Persson, Per Pettersson-Lidbom, Federico Revelli, Christian Riis, Elisabeth Schulte-Runne, Christoph Vanberg, Galina Zudenkova and seminar participants at universities in Freiburg, Heidelberg, Konstanz, Mannheim, Norwegian Business School, the Silvaplane Workshop on Political Economy, the Conference on Evaluation of Political Reforms in Mannheim, the European Public Choice Society Meeting, the CESifo Conference on Public Sector Economics, the IIPF and the EEA Annual Meeting for helpful suggestions. We are very grateful to Greg Shaw, Dave Andersen, and Paul Ehmann for sharing some of their data, to Shelly Arsenault for providing supplemental material on state waivers and the administrators of the Welfare Rules Database (WRD) at the Urban Institute for answering our questions. Jan Scherer, Michael Hellwig, and Robert Aue provided outstanding research assistance. We are responsible for all remaining errors.

<sup>†</sup>Department of Economics, University of Mannheim, andreas.bernecker@absolventum.uni-mannheim.de

<sup>‡</sup>Department of Economics, University of Mannheim, CESifo, pierre.boyer@uni-mannheim.de

<sup>§</sup>Department of Economics, University of Heidelberg, IZA and CESifo, christina.gathmann@awi.uni-heidelberg.de (*corresponding author*)

# 1 Introduction

When President Clinton signed the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) in August of 1996, it marked the most important reform of US welfare policy since the New Deal. PRWORA abolished the entitlement to cash assistance for poor families, the *Aid to Families with Dependent Children* (AFDC). The new *Temporary Aid to Needy Families* (TANF) defined welfare benefits as temporary assistance rather than as an entitlement to eligible families. At the same time, PRWORA decentralized the authority to design welfare policies to the individual states (see, for example, Grogger and Karoly (2005), Moffitt (2008)). After 1996, policy-makers in the states were largely autonomous to choose eligibility criteria for benefit receipt, the requirements recipients had to fulfill while on welfare, and the type of sanctions imposed if recipients did not comply with these requirements.

A vast literature has analyzed the consequences of the 1996 welfare reform for caseloads, employment, earnings, poverty rates, living arrangements, marriage patterns and other socio-economic outcomes (see, for example, Bitler and Hoynes (2010) and Meyer and Sullivan (2008) for some recent evidence; and Blank (2002), Grogger and Karoly (2005), Moffitt (2001), and Moffitt (2002b) for excellent surveys).

In sharp contrast to the literature on the effects of the welfare reform, very little is known about the political economy of welfare reform in the US states.<sup>1</sup> Many welfare rules had never been tried before and most states had little experience how certain welfare rules would work in their local jurisdiction. If little is known about the effectiveness of certain policies, there are potentially large gains from learning through experimentation and possibly large political costs of failure. Yet, how much did states take advantage of the opportunity to experiment and learn before and after PRWORA? And what determined whether states actually experimented with new policies and possibly reversed welfare rules later on?

Our analysis provides a first answer to these important questions. To guide our

---

<sup>1</sup>In contrast, several detailed studies track the political process leading up to reform at the federal level in 1996 (Reintsma (2007), Weaver (2000), or Haskins (2006)).

empirical investigation, we outline a political agency model of policy experimentation, inspired by Majumdar and Mukand (2004). In the model, voters are uncertain about the best welfare policy but politicians may have some information depending on their experience. Politicians care about voters' welfare as well as their chances of reelection. During her term in office an incumbent can decide whether to stick to the known status quo policy or experiment with a new, uncertain policy. The experiment is ex-ante welfare-enhancing but may yield its benefits in the short- or long-run. After implementing the experiment, the incumbent observes its short-term outcome. If the experiment has no short-term benefit, it is either because the experiment was implemented in the wrong state of the world or because of bad luck. The politician can now decide whether to stick with the policy or revert back to the status quo. Prior to the next election, the final outcome of the policy choices is observed by both politician and voters.

A politician, uncertain whether the experiment is beneficial, weighs the (risky) reputational gains from an experiment against her safe initial reputation. The higher her initial reputation (and hence here reelection chances ex-ante), the more an incumbent has to lose if the policy experiment fails. Hence, reputational concerns makes the politician hesitant to engage in the risky experiment in the first place; as a result, there is too little policy experimentation relative to the first-best.

This reputation effect should be weaker for governors who put a higher weight on voter welfare and worry less about their reelection chances. The reason is that the experiment increases ex-ante welfare. Hence, the less a politician worries about reelection, the more likely she engages in experimentation. The model also has something to say when a policy experiment might be reversed. If an experiment fails in the short-term, it is either the wrong policy for the state of the world or bad luck. A politician with little information whether the experiment is beneficial in the current state of the world is more likely to have made a mistake when experimenting than a politician who knows the state of the world. Yet, if the potential benefit of an experiment is large, reversal becomes less likely. The reason is that successful experimentation yields a large welfare gains and thus increases the incumbent's willingness to gamble with voters' welfare.

To test our predictions in the context of the US welfare reform, we collect a new dataset on welfare policy rules both before and after the PRWORA reform. Our analysis thus spans both the period of welfare waivers states adopted under the old AFDC program as well as the policy changes implemented after the federal reform in 1996. We complement our panel of policy rules with detailed information on the political structure and socioeconomic conditions in each state.

Our results provide strong support for the idea that reelection concerns influence policy experimentation. A high initial reputation of the politician reduces policy experimentation both during the waiver and TANF period, especially among governors who can be reelected. The influence of reelection concerns on experimentation is economically sizable: an increase by an interquartile range in initial reputation reduces experimentation by 32-35%. Furthermore, governors with relatively little experience are more likely to experiment when they are less concerned about reelection. Again, the effect is sizable: lame ducks with relatively little experience are 43% more likely to experiment. We also find that governors with low experience are more likely to revert an experiment, but less likely when the potential gains of experimentation appear high.

A number of robustness checks bolster our findings that reelection concerns matter for policy experimentation. Most importantly, one might expect that ideology influences the decision to experiment as Republicans pushed for the idea of workfare instead of welfare. Surprisingly, we find that ideology of the governor or voters has little influence on experimentation. Our results are also robust to controlling for the influence of the legislature as measured by party composition, legislative polarization or the contestedness of state elections. Politicians trying to decide whether to implement a policy might look to their neighbors for inspiration or engage in welfare competition. We construct several measures of “neighbors” in terms of borders, size and ideology but find little evidence that experimentation is explained by cross-state spillovers. Similarly, incentives or pressures to experiment may also depend on local economic or social condition. A state’s socio-demographic structure indeed matters, but has little effect on our main results. In sum, experimentation with welfare waivers during the AFDC program and welfare rules during

the TANF program both support the idea that reputation concerns are important to understand policy innovations during the US welfare reform.

The paper is structured as follows. The next section discusses the related literature while Section 3 provides background information about the US welfare reform. In Section 4, we introduce our political agency model and derive the empirical predictions. We describe the data sources and empirical strategy in Section 5. Section 6 presents the main results and the reports robustness checks of our evidence. The last section concludes.

## 2 Related Literature

Our analysis is closely related to the idea of a laboratory federalism where states may engage in experimentation to learn about the best policy in a world of imperfect information (see e.g. Oates (1999) for a survey). We show that reelection concerns and accountability may reduce incentives to experiment and possibly reverse an experiment later on. We also test for potential spillover effects, i.e. the idea that states may learn from and imitate neighboring states in the same federation.

As such, we also contribute to a sizable political science literature on policy innovation and diffusion such as the adoption of state lotteries or labor market regulation (e.g. Walker (1969), Cnudde and McCrone (1969), Allen and Clark (1981), Berry and Berry (1990), Berry and Berry (1992), or Boehmke and Skinner (2012); see Karch (2007) and Berry and Berry (2007) for surveys). Our theoretical and empirical focus however, is on the political economy of decision-making during the US welfare reform, considered one of the major social reforms in the US since the New Deal. In particular, we seek to understand the role of reelection concerns for policy experimentation and reversals.<sup>2</sup>

Our analysis also contributes to a growing empirical literature on the role of electoral concerns. Most papers study the link between gubernatorial term limits and tax or spending decisions (see e.g. Besley and Case (1995a), Besley and Case (2003) and List

---

<sup>2</sup>One exception is Lieberman and Shaw (2000) who analyze the choice of welfare rules as a function of local conditions. Also, Soss et al. (2001) study how an array of factors such as ideology, racial composition or electoral competition, affect the policy choices of states in the TANF program.

and Sturm (2006)). Alt and Rose (2011) find evidence that term limits exhibit both an accountability effect (where governors who can be reelected perform better than term-limited governors) and a competence effect (experience in office improves performance) where performance is measured by high economic growth and low spending, taxes and borrowing costs. A recent study by Ferraz and Finan (2011) demonstrate that electoral accountability reduces corruption of local politicians in Brazil. We focus instead on policy experimentation and learning from experimentation. In our case, reputation effects may result in too little experimentation if the politician is uncertain about the benefits of the experiment and puts a high weight on reelection.

Our theoretical setup is related to a large literature on reputational concerns following the seminal work of Holmström (Holmström (1982), Holmström (1999)). An early application to the political arena is the political agency model by Rogoff (1990). Closer to our setting are models where voters and politicians are both uncertain about the mapping from chosen policies to realized outcomes (Harrington (1993), Majumdar and Mukand (2004), Fu and Li (2014), or Willems (2013)).<sup>3</sup> As in our setting, politicians are trading off the potential welfare and electoral gains when undertaking (and continuing) a policy reform against the reputation loss that a failure would deliver. The focus in our paper is on bringing some predictions of these models to the data. Our analysis is the first to test the implication of reputational concerns for incentives to innovate and learn in the political arena.

Our study is also closely related to the literature on policy persistence which tackles the question why certain policies persist even if they are known to be a failure.<sup>4</sup> Our focus on the reputational costs of policy reversals as one explanation why policies might persist

---

<sup>3</sup>Recent papers by Callander (2011a), Callander (2011b), and Callander and Hummel (2014) model policy experimentation with a richer policy space: their “world” is not restricted to a two-state world and policies in these papers can take more values than the binary policy choice (reform or not) adopted in most of the earlier literature. Callander and Harstad (2013) present a model where heterogeneous districts choose both whether to experiment and the policies to experiment with. Reelection concerns and reputation of the policy makers play no role in their analysis. Instead, we keep a simple setup with two states of the world and a binary decision in order to study the consequences of reputational concerns for experimentation and reversals.

<sup>4</sup>Potential mechanisms stressed in the previous literature are individual-specific uncertainty about the winners and losers of a reform (Fernandez and Rodrik (1991), Jain and Mukand (2003), Ciccone (2004)), a war of attrition between politicians (Alesina and Drazen (1991)), or vested interests benefiting from the status quo (Coate and Morris (1999)).

differs from potential mechanisms highlighted in previous papers. Most importantly, we provide an empirical test whether reputation matters for policy persistence in the context of the US welfare reform.

## 3 The 1996 Welfare Reform in the United States

### 3.1 AFDC Program and Welfare Waivers

Since the New Deal, Aid to Families with Dependent Children (AFDC) provided financial assistance to needy children lacking parental care or support (see Grogger and Karoly (2005), Moffitt (2008), or Bitler and Hoynes (2010)). The program was jointly administered by the federal and state governments though most of the eligibility criteria and provisions were determined by the federal government.<sup>5</sup> States could choose the level of monthly benefits; otherwise, they had little room to design or shape federal welfare policy.<sup>6</sup>

Starting in the early 1960s, state governments obtained more autonomy to experiment with their own welfare rules. Since 1962, Section 1115 of the Social Security Act allowed the Secretary of Health and Human Services to waive federal AFDC rules and regulations. States could petition the U.S. Department of Health and Human Services (DHHS) to implement other welfare rules in temporary pilot projects.<sup>7</sup>

To apply for a waiver, state governments submitted detailed requests which rules and program elements they planned to modify and the new rules and regulations to be implemented instead.<sup>8</sup> The proposed waiver provisions were then reviewed, and

---

<sup>5</sup>A separate program for jobless two-parent families (AFDC-UP) was established in 1961, but single-parent families remained the primary beneficiaries of the AFDC program prior to 1996.

<sup>6</sup>For a family of three with no other income, for instance, monthly benefits in 1995 varied from \$119 in Mississippi to \$720 in Alaska. States could also implement a few eligibility rule like “fit parent” or “suitable home” provisions in order to limit payments to families with unsatisfactory behavior.

<sup>7</sup>These experiments (which were often implemented in a few local areas) had to be cost-neutral and to be evaluated scientifically.

<sup>8</sup>Most waivers implemented multiple changes to welfare rules simultaneously (see Crouse (1999) and Koerper (1996) for detailed descriptions of the state waivers). Most popular were experiments to impose stricter work requirements (to increase work participation among recipients), impose time limits on benefit receipt, sanctions (for benefit units failing to meet work requirements), and family caps (eliminating or reducing the benefit increase if a child is conceived during welfare receipt).

sometimes altered, by the federal offices with jurisdiction over the various aspects of each state's proposal. The Secretary of Health and Human Services made the final decision whether to approve a waiver, request changes or deny it. In most cases, the waiver was approved as is or after some changes.

Welfare waivers were rare until the late 1980s when AFDC caseloads approached almost 4 millions families or around 11 million recipients nationwide (see Figure 1 for the evolution of per capita caseloads). Waivers became much more popular during the Reagan years when federal aid for AFDC declined and government funds tightened during the 1989-1992 recession. Between 1987 and 1992, 15 waiver applications in 14 states were approved during the Reagan administration and another 15 applications from 12 states were approved during the Bush administration (see Harvey et al. (2000)). As reform efforts accelerated during the first term of the Clinton administration, the federal government approved 83 waivers. In total, all but five states received approval for one or more waivers (see Table A1 in the appendix for a list of the waiver applications by state).

In many cases, the initiative to experiment with welfare waivers during the 1980s and early 1990s came from state governors. Wisconsin's governor Tommy Thompson, a leading figure in the welfare reform process, provides a good example. He made welfare reform a top priority in his campaign for governor as early as 1986 (Mead (2004), Kaplan (2000)). When elected in 1987, Thompson immediately created a task committee to reform the existing AFDC system. Between 1987 and 1996, the Thompson administration applied for three waivers in 1988, 1992 and 1993. In 1988, Wisconsin was the first state which conditioned a household's receipt of benefits on the school attendance of its teen children.

Governor Thompson, by using his line item veto power, implemented the new Wisconsin Works, better known as W-2, in 1993. The W-2 program, with its emphasis on labor market participation, stricter work requirements and harsh sanctions for noncompliance, became a blueprint for the federal TANF reform in 1996. Governors in Delaware, Michigan, Minnesota, Nebraska and Ohio played similar defining roles for the speed and direction of welfare reform (see Weissert (2000) or Winston (2002) for case

studies).

### 3.2 The Introduction of TANF

The passage of the Personal Responsibility and Work Opportunity Reconciliation Act (PROWRA) in August of 1996 abolished AFDC and replaced it with the *Temporary Assistance to Needy Families* (TANF) program.<sup>9</sup> The PRWORA reform reshaped the landscape of US welfare policy along at least three dimensions. Most importantly, the reform decentralized most of the authority to design welfare programs to the state level. After six decades of the federal AFDC program, state governments were free to choose welfare rules, however generous or restrictive, bound only by a few federal guidelines.<sup>10</sup> Waiver provisions remained in place until the waiver expired or were replaced by new statewide TANF rules. Because states had to decide on many new policy dimensions, policy rules changed even after the adoption of TANF at the state level.

The second major element of the reform was the explicit goal to reduce welfare caseloads and boost work participation (i.e. workfare instead of welfare). By setting federal time limits to benefit receipt, for example, TANF introduced the idea that state support was a temporary measure rather than a permanent source of income.<sup>11</sup> The work requirements in the TANF legislation were much stricter than under AFDC, and focused on active participation in the labor market rather than training and education. Recipients who failed to comply with the new work requirements, also faced more rigorous sanctions than under AFDC.

The third novelty was that federal funding for the TANF program was consolidated

---

<sup>9</sup>State governors played a prominent role in pushing and keeping federal welfare reform on the political agenda as well. The National Governors' Association (NGA) was influential in lobbying for welfare reform at the federal level (see Weaver (2000)). Fourteen governors testified in welfare hearings of the 104th Congress compared to only three state representatives (Winston (2002)).

<sup>10</sup>Between 1996 and 1998, states then replaced their existing AFDC programs with new state TANF programs. Implementation dates ranged from September 1996 for Massachusetts, Michigan, and Vermont to January 1998 for California.

<sup>11</sup>The federal government sets a lifetime time limit for benefit receipt of 60 months. States that wish to offer longer time limits (or no time limit at all) need to use state funds to support recipients for more than 60 months. States could also decide to implement shorter lifetime time limits on benefit receipts. We discuss the specific welfare rules that are part of the empirical analysis in more detail in the data section and data appendix.

into a single block grant. Any additional funds, from rising caseloads in economic recessions or through less restrictive eligibility criteria, had to be financed by the state alone.<sup>12</sup> PRWORA also imposed financial penalties, in the form of percentage reductions in the block grant, on states that did not comply with the federally mandated work participation rates of recipients or caseload sizes.

As a consequence of the new financing scheme, states had strong incentives to reduce welfare caseloads as to avoid additional burden on the state budget. These incentives are also reflected in the actual numbers: for example, caseloads decreased nationwide by more than 56% between 1994 and 2000 (see Bitler and Hoynes (2010)).<sup>13</sup> Despite the common focus on workfare and reducing welfare dependence, the extent and speed of activity during the waiver and TANF period differed a lot across states. We next provide a framework to shed light on the determinants of policy experimentation during the US welfare reform.

## 4 Theoretical Framework

To study the decision to experiment with (and possibly reverse) welfare policies in the US states, we outline a simple model inspired by Majumdar and Mukand (2004). Our main goal is to derive predictions which we can then take to the data. Given the dominant role played by state governors both during the waiver period and the TANF period, we focus on a single political decision-maker who is concerned about voters' welfare and reelection. To simplify the exposition, all proofs are relegated to the Appendix.

**Policies.** Consider an elected politician who has an opportunity to enact a reform  $a_N$ . The incumbent needs to decide whether to continue with the safe status quo policy  $a_S$

---

<sup>12</sup>Each state's block grant was funded at the annual spending level corresponding to fiscal years 1992 to 1995. States were also required to contribute substantially to other program components that were replaced by TANF like subsidized child care. Under the old AFDC program, in contrast, a matching grant by the federal government co-financed state and local contributions. The costs of additional caseloads in a state were therefore shared by the state and the federal government.

<sup>13</sup>The booming US economy of the 1990s is responsible for about a third of the caseload reduction; the remainder may be attributed to states' incentives to tighten access to welfare benefits (see, for example, Council of Economic Advisors (1999) and Fang and Keane (2004)).

or experiment with the new policy  $a_N$ . Both policies affect overall welfare which is publicly observable. Different policies are appropriate for different environments; hence, the success of the new policy  $a_N$  is contingent on the underlying state of the world. The state of the world may be one of two types denoted by  $S$  and  $N$ , with  $a_S$  and  $a_N$  being the appropriate policies for the two environments respectively. If the underlying state of the world is  $S$ , then enacting the new policy  $a_N$  causes a net loss in welfare. The state  $N$  (resp.  $S$ ) occurs with probability  $p$  (resp.  $1 - p$ ).

The status quo policy  $a_S$  is assumed to deliver a fixed welfare, normalized to 0, independently of the state of the world. No benefits from the new policy occur in state  $S$ . In state  $N$ , the new policy initiative delivers a benefit  $\Delta$ . However, this benefit occurs either in the short-term or in the long-term. The benefit realizes in the short-term (resp. in the long-run) with probability  $q$  (resp.  $1 - q$ ).

The total cost of enacting and continuing the new policy initiative is  $c$  which is incurred independently of the success of the reform. This cost may represent the additional training of the bureaucracy to effectively administer the new policy or other costs of implementation. The total cost  $c$  consists of short- and long-run cost so that  $\frac{c}{2}$  is incurred when the reform is implemented and again if the reform is maintained in the long-term. If the reform is reverted back to  $a_S$  instead, then the long-term cost is not incurred. However, the short-term benefit cannot be kept in the long-term if the reform is reversed.

We make the following assumptions:

$$p\Delta - c > 0, \tag{A1}$$

$$pq\Delta - \frac{c}{2}(1 + pq) > 0, \tag{A2}$$

and

$$\beta\Delta - \frac{c}{2} < 0, \tag{A3}$$

where  $\beta = \frac{p(1-q)}{1-pq} < p$  is the probability that the state of the world is  $N$  conditional on a short-term failure of the reform.

Assumption (A1) implies that a reform that is continued has positive expected welfare gains from an ex-ante perspective. Assumption (A2) implies that undertaking the reform and reverting back to the status quo in the face of short-term failure has positive expected benefit from an ex-ante perspective. Finally, assumption (A3) implies that, if the benefit does not realize in the short-term, the expected future benefit from continuing with the reform is negative.<sup>14</sup>

**Timing.** The game has three periods,  $T = \{0, 1, 2\}$ . At the beginning of the game ( $T = 0$ ), the incumbent politician has the opportunity to change the existing policy  $a_S$ . She receives a private signal about the appropriateness of the policies, i.e. the state of the world, and then faces the choice of either maintaining the status quo policy  $a_S$  or enacting the new policy  $a_N$ . If she decides to maintain the status quo, then the ex-post welfare is 0. On the other hand, if the new policy is enacted, a cost  $\frac{c}{2}$  is initially incurred. Then, the politician learns about the realization of short-term benefit from the reform at the beginning of  $T = 1$ . If the new policy turns out to be successful, the politician knows that the state is  $N$ , and therefore  $a_N$  is the appropriate policy. This implies that the benefit  $\Delta$  is secured by continuing the reform and incurring the long-term cost  $\frac{c}{2}$  at the beginning of period  $T = 2$ . However, in the case where the benefit does not realize in the short-term, it is not clear whether this failure is due to bad luck or because the policy is not appropriate. The politician now has to decide whether to continue with the policy initiative  $a_N$  or to revert back to the initial status quo  $a_S$ . If she revert back to  $a_S$ , ex-post welfare is  $-\frac{c}{2}$ . If she continues with  $a_N$ , the society again incurs a cost  $\frac{c}{2}$  and gets the long-run benefit only if the state of the world is  $N$ . The ex-post welfare is observed before the next election that takes place at the end of period  $T = 2$ , corresponding to the end of the incumbent's term in office.

**Politicians.** The economy has many politicians that differ in their capability to assess the appropriateness of policies for the economy, and can be either of high quality or of

---

<sup>14</sup>We implicitly assume that the parameters of the model take values such that assumptions (A1), (A2), and (A3) hold simultaneously. This is the case for a large range of parameter values.

low quality. The incumbent politician knows her type. For simplicity, we assume that a high-quality incumbent knows the state of the world. On the other hand, a low-quality politician does not know anything *ex ante*, i.e. she only knows that the state of the world  $N$  is realized with probability  $p$ , and that in this state the benefit realizes in period 1 (resp. period 2) with probability  $q$  (resp.  $1 - q$ ).

The incumbent politician cares about both welfare of her constituency as well as her own future electoral prospects. The incumbent's objective function is

$$\gamma(\text{Welfare}) + (1 - \gamma)(\text{Probability of Reelection}), \quad (1)$$

where  $\gamma \in [0, 1]$  is the relative weight that she puts on welfare.

Finally, at the beginning of the game ( $T = 0$ ) the incumbent has earned a reputation among the electorate represented by  $\lambda \in [0, 1]$ . We assume that this reputation does not depend on future decisions taken by the incumbent and remains constant until the next election.<sup>15</sup>

**Voters and Reelection Rule.** The electorate consists of a representative voter that chooses either to reelect or throw the incumbent out of office at the end of the term  $T = 2$ . The voter cares only about *ex-post* welfare. We assume that the voter does not know that there are high- or low-quality politicians so that he does not think that the incumbent and the (large number of) potential challengers differ in competence. However, the voter may take into account the reputation of the politician when deciding to reelect the incumbent.

Assuming the representative voter does not know that politicians differ in abilities simplifies the analysis considerably. Under this assumption, there is no signaling game in the policy choice of the incumbent: both the challenger and the incumbent are perceived to be identical by the voter except in the case where the incumbent's reputation is positive.<sup>16</sup>

---

<sup>15</sup>See Majumdar and Mukand (2004) for an analysis where the reputation of the incumbent changes with the decision to undertake reforms.

<sup>16</sup>See Majumdar and Mukand (2004) for an analysis of the signaling game where the representative

We assume that the representative voter follows a simple reelection rule: reelect the incumbent with probability 1 if ex-post welfare is positive; never reelect the incumbent if ex-post welfare is negative; and reelect the incumbent with probability  $\lambda$  if ex-post welfare is 0. This reelection rule is in the spirit of agency models where voters punish the incumbent for bad performances (see Persson and Tabellini (2000)).

## 4.1 Policy Choices of a High-quality Politician

We first analyze the decision of a high-quality incumbent.

**Proposition 1.** *For any relative welfare weight  $\gamma$  and any reputation  $\lambda$ , a high-quality incumbent always undertakes a reform if and only if the state is  $N$ . In addition, a reform undertaken is never reverted back to the status quo.*

A high-quality politician knows the state of the world. If the state of the world is  $S$ , since there are no welfare gains and therefore no electoral gains from a reform, she will never experiment. Proposition 1 shows that a high-quality incumbent has her objective fully aligned with the interest of the representative voter: there is no conflict of interest between maximizing voter welfare and her reelection chances.

## 4.2 Policy Choices of a Low-quality Politician

We now show that the decision-making of a low-quality incumbent may exhibit a conflict between maximizing welfare and reelection chances. In order to disentangle the different forces, we start by developing two polar cases: the one of a purely welfare-maximizing politician ( $\gamma = 1$ ) and of a purely office-motivated politician ( $\gamma = 0$ ).

**Welfare-maximizing Politician.** The following proposition presents the case where the policy choices of the incumbent is driven only by welfare considerations.

---

voter makes inferences on incumbent quality based on the sequence of policy choices and the realized outcomes. The predictions they derive are similar to the ones in the simplified model presented here.

**Proposition 2.** *A low-quality incumbent who only cares about welfare ( $\gamma = 1$ ) always implements the reform, and always reverts back to the status quo if the reform does not deliver the short-term benefit.*

The decision-making policy described in Proposition 2 is the optimal one from a welfare perspective under the veil of ignorance. Indeed, for a low-quality politician who does not know the state of the world, the reform is welfare maximizing from an ex-ante perspective. However, it is welfare-maximizing to revert back to  $a_S$  if the reform fails in the short-run since the expected future benefit from the reform is then negative.

**Office-motivated Politician.** The second benchmark case presents the policy choice of a purely office-motivated politician.

**Proposition 3.** *A low-quality incumbent who is purely office-motivated ( $\gamma = 0$ ) implements the reform if and only if*

$$p \geq \lambda, \tag{2}$$

*and never reverts back to the status quo when the reform is undertaken.*

Proposition 3 shows that it is the reelection motive of the low-quality politician that leads to policy persistence in the face of short-term failure. A low-quality politician might therefore not maximize voter welfare for two reasons. First, equation (2) implies that the politician’s initial reputation can conflict with the decision to experiment with profitable reforms. Second, in the face of short-term failure a purely office-motivated politician does not take into account the expected welfare loss of continuing with the reform and “gambles” for reelection with the welfare of the electorate.

**Corollary 1.** *Purely office-motivated politicians experiment less when they have higher initial reputation among the electorate. In particular, there exists a cut-off  $\lambda^* := p$  such that low-quality politicians try the reform if and only if their reputation is below  $\lambda^*$ .*

Corollary 1 characterizes the cut-off  $\lambda^*$  which determines the reputation level that leads to inefficient policy experimentation. Indeed, following Proposition 2, for an

uninformed politician always trying the reform is welfare-maximizing from an ex-ante perspective. However, a low-quality politician with a high enough initial reputation, i.e. above  $\lambda^*$ , will never experiment with the reform. As a consequence, reputation concerns result in too little policy experimentation from a welfare perspective.

**Intermediate Politician.** We now turn to the general case where the low-quality politician is concerned with both welfare and her reelection chances ( $\gamma \in [0, 1]$ ). We focus on the decision to revert an experiment.<sup>17</sup>

**Proposition 4.** *An incumbent politician with low-quality always reverts back to the status quo in face of a short-term failure of the reform if and only if  $\gamma \geq \gamma^*$ , where*

$$\gamma^* := \frac{1}{1 + \frac{c}{2\beta} - \Delta}. \quad (3)$$

A low-quality politician who observes the short-term failure of her reform does not know whether the reform will be successful in the long-term. This effect pushes her to revert back to the status quo in order to increase expected welfare as described in Proposition 2. However, reelection concerns pull her towards sticking with the reform as in Proposition 3. Proposition 4 disentangles these two conflicting forces which affect the decision to revert back to the status quo or not. Equation (3) defines a threshold level  $\gamma^*$  so that for values of relative welfare weights above the threshold, the low-quality incumbent reverts the reform as welfare considerations outweigh reelection considerations.

**Corollary 2.** *A low-quality incumbent is more likely to revert in the face of short-term failure of the reform when  $\gamma$  increases and  $\Delta$  decreases.*

Corollary 2 shows how the likelihood of a policy reversal is related to model parameters which leads to a better alignment of the objectives of the low-quality incumbent and the representative voter, as described in Proposition 2. High-quality politicians never revert

---

<sup>17</sup>For completeness we also treat the decision to experiment in Appendix A.2. We relegate this case to the Appendix because it does not yield any important insights for our empirical analysis beyond those discussed in Propositions 2 and 3.

an experiment (Proposition 1). In contrast, low-quality politicians make their reversal decision dependent on the welfare gains and costs of the experiment (Corollary 2).

### 4.3 Empirical Predictions

Our results above characterize how reelection concerns affect policy experimentation (and reversals) and provide four testable implications.

The first empirical prediction is that the higher the initial reputation of politicians, the less likely they are to experiment. Politicians who are purely office-motivated are less likely to reform because they weigh the risky reputational gains from an experiment against their safe initial reputation (Proposition 3 and Corollary 1).

The second prediction is that this reputation effect is weaker for politicians who put a higher weight on voter welfare. Since the experiment increases welfare ex-ante, politicians who are less concerned about reelection than about welfare, experiment more (Proposition 2 as well as Proposition 5 and Corollary 3 in the Appendix for intermediate cases).

The third prediction is that a low-quality politician may be more or less likely to experiment than a high-quality politician. The high-quality politician experiments when the state of the world is  $N$  which occurs with probability  $p$  (Proposition 1). The low-quality politician is more likely to experiment if  $\gamma = 1$  (Proposition 2) but less likely if  $\gamma = 0$  and  $\lambda \geq p$  (Proposition 3). Yet, a low-quality politician is more likely to experiment when the weight she puts on voter welfare increases (Corollary 3 in the Appendix).

Our final prediction relates to policy reversals: low-quality politicians are more likely to revert an experiment than high-quality ones. The reason is that the low-quality governor does not know whether the experiment is appropriate for the current state of the world; hence, she is more likely to have made a mistake. The likelihood of a policy reversal by a low-quality politician decreases with the potential welfare gains from the experiment (Proposition 4 and Corollary 2).

## 5 Data Sources and Empirical Strategy

### 5.1 Data Sources

To test our predictions, we collect a comprehensive dataset of welfare policy rules, political conditions and socio-economic characteristics for all US states between 1978 and 2010. Here, we introduce our key variables. The appendix provides more details about the data sources and the construction of all variables used in the empirical analysis.

**Policy experimentation and reversals.** We measure policy experimentation during the AFDC period by the waivers applications at the Secretary of Health and Human Services (from Koerper (1996), Crouse (1999), and Lieberman and Shaw (2000)). To reduce any bias by the federal review process, we include waiver applications that have been approved and implemented, but also those that were denied by the federal agency or withdrawn by the state. Our measure also counts waiver applications irrespective of whether the state planned to implement it only in specific counties or the state as a whole. Table A1 shows the distribution of waivers across states. Most states only apply for one waiver within a year while twenty states applied for two waivers within the same year. Our measure of policy experimentation prior to 1996 is then a simple count variable of the number of waivers a state applied for in a year. If a state did not apply for a waiver, the count variable is set to zero. On average, 0.2 waivers per year were applied for during the 1978-1996 period.

Defining an experiment is much more difficult for the post-1996 period. As a first step, we collect welfare policy rules in each state and year from the Welfare Rules Database of the Urban Institute (Urban Institute (2012)). The Welfare Rules Database provides the most comprehensive and up-to-date information on welfare rules in all fifty states for the post-1996 period. For our analysis, we focus on a subset of policy rules that were at the center of the public and academic debate (see the discussion in Bitler and Hoynes (2010); Fang and Keane (2004); or Grogger and Karoly (2005) for example) surrounding the US welfare reform: (1) Whether the state has a family cap, which limits the additional

benefits if a child was conceived and born while the parent is on welfare. (2) The main rules governing time limits for benefit receipt. (3) What work requirements adults have to fulfill in order to remain eligible for benefit receipt. And (4) the primary sanctions if a household does not fulfill the work requirements. The data appendix and Table A2 provide more detailed explanations about the individual policy rules.<sup>18</sup>

To illustrate how we measure an experiment or reversal, take the example of family caps which did not exist under the old AFDC program. Between 1996 and 2010, twenty-four states adopted a family cap, eight of these twenty-four states later abolished the family cap previously adopted. We then code a policy experiment if a state first adopts a family cap after 1996 and we code a policy reversal if a state (that had adopted a family cap until 1996 or later) abolishes it in any year between 1996 and 2010.

We code policy reversals for other policy rules in a similar way: a tightening of a policy rule in a specific year is coded as an experiment, abolishing that restriction in a later year is coded as a policy reversal. Both variables are zero if no policy experimentation or reversal took place in a state in that year (see Figure 3 for the evolution of experimentation and reversals).<sup>19</sup>

**Electoral incentives and potential gains.** Our main independent variables are the governor's initial reputation (an empirical counterpart to  $\lambda$ ), the importance of reelection

---

<sup>18</sup>As a sensitivity check, we construct alternative measures using different sets of policy rules (see Table A5). One alternative measure restricts experiments to the first change in a rule but ignores subsequent rule changes in the same direction. The second alternative analyzes an extended set of welfare rules which also includes: earnings disregards for benefit calculation; whether a state requires households to look for a job prior to applying for benefits (mandatory job search); and whether a state offers temporary payments to families if they do not apply for welfare benefits (so-called diversion payments). As shown below, we obtain comparable results using these alternative measures of policy experimentation.

<sup>19</sup>Note that we can measure the adoption and abolishing of specific policy rules unambiguously only for the years after 1996 (when covered by the Welfare Rules Database of the Urban Institute). Changes in policy rules during the waiver period are not captured systematically in the database because the information contained in the waiver application cannot be merged to the Welfare Rules Database in a consistent manner. If a state both adopted and abolished a family cap prior to 1996, we will not capture this in our measure of policy reversal which is only available after 1996. Anecdotal evidence however suggests that policy reversals were rare before 1996 as many waivers only expired after the federal TANF reform. Similarly, we might have a bias in our post-1996 policy experimentation measure as well. Suppose a state experiments with a policy rule, say a family cap, using a waiver, but did not have a family cap in 1996 (as coded in the Welfare Rules Database). If the waiver provisions are still in place in 1996, any changes will be captured by our experimentation measure; if in contrast, the waiver expired before 1996 and the state would again adopt a family cap after 1996, we would code it as an experiment (even though the state has tried this experiment before in the waiver period).

concerns (a proxy for  $\gamma$ ), a proxy for the competence of the governor; and finally, some measure indicating the potential gains from experimentation (a proxy for  $\Delta$ ). All of these concepts are difficult to measure as they are ultimately unobservable to the researcher.

We use the vote margin in the past gubernatorial race to proxy the governor's initial reputation. Governors which voters believe to be capable are likely to be elected with a larger majority than governors that just barely obtain a majority. Below, we return to the issue how to distinguish the reputation effect from other potential interpretations of the vote margin.

To capture reelection concerns among governors, we use information on whether a state governor faces a binding term limit in her current term in office, i.e. whether she is a lame duck. Lame ducks might still have some electoral incentives, for example, because they plan to run for another office at the federal level. All we need for our analysis is that lame ducks care somewhat less about their electoral prospects in their state than governors who can and want to remain in office.

To measure the competence of the governor, we use the years a governor has spent in all political offices before being elected as governor (see also Smart and Sturm (2013) for a similar approach). Cumulative years in political offices will be a good indicator of the governor's ability to recognize what is appropriate in the current state of the world if elections select the most able candidate or candidates acquire competence in policy-making over time. While a governor's tenure in the current office is in principle observable to voters, voters are unlikely to know exactly how much political experience a governor has accumulated prior to being governor.<sup>20</sup>

The most difficult task is to find a proxy for the potential gain from a policy experiment. An important motivation for state governors to apply for a waiver was

---

<sup>20</sup>A few recent studies use formal educational qualifications as a proxy for quality (see Galasso and Nannicini (2011) and Gagliarducci and Nannicini (2013) for Italian politicians or Kotakorpi and Poutvaara (2011) for Finnish members of parliament). In our analysis, educational qualifications were never found to be statistically or economically relevant for policy innovations. One possible explanation is that most governors in our sample are highly skilled: only 6% do not have a college degree or more, while almost 50% have a master or other postgraduate degree. In contrast, the average local politician in Italy only has a high-school degree; in the Finnish parliament, about 20-30% of its members have a masters degree or more. Hence, formal educational credentials might be more a prerequisite for becoming governor rather than a signal of quality. Furthermore, voters are probably more aware of the educational credentials than of the governor's experience in all prior political offices.

to reduce caseloads and hence, welfare spending. The incentive to reduce state spending became even stronger during the TANF period because states have to cover all welfare expenditures exceeding the federal block grant.<sup>21</sup> As a broad proxy for the potential gain from experimentation, we use a state's AFDC/TANF spending (lagged by two years to reduce feedback effects).

**Control variables.** To control for other potential influences on the decisions to experiment and reverse a policy, we collect information on ideology. We obtain the governor's party and a measure of governor ideology (based on Berry et al. (1998)) which takes into account that governors belonging to the same party might have different ideologies. The ideology measure varies between 0 and 100 where larger values represent a more liberal attitude.

To control for legislative influences, we use measures of party composition and political polarization. Polarization is measured as the Democratic seat share in the lower (upper) chamber calculated as deviation from 50%. We further code whether the government in the state is politically divided between Democrats and Republicans; and a measure of competitiveness in state legislative elections using the Holbrook-Van Dunk index (see Holbrook and Van Dunk (1993)).

To control for the demand for welfare, we collect information on personal income per capita, the size of the population, the share of the Black population in a state, the age structure, the immigrant population, the share of unmarried births and income inequality (measured as household income at the 90th over the 10th percentile). Finally, states might also learn from or imitate other states. To analyze these cross-state spillover effects, we define geographic neighbors (states sharing a boundary), states similar in population size (based on population in 1978) and ideological neighbors (with a similar voting record in the 1978 presidential election). Table A3 contains summary statistics of all variables used in our analysis over the whole period (1978-2010).

---

<sup>21</sup>States faced also reductions in the block grant if they did not reach the federally mandated employment rates among benefit recipients.

## 5.2 Empirical Strategy

Our basic empirical approach is a simple difference-in-difference specification (see Besley and Case (1995a); or List and Sturm (2006) for similar approaches). To test our first prediction, we estimate the following model:

$$Experiment_{st} = \alpha_1 Reputation_{st} + \pi_1 LD_{st} * Reputation_{st} + \phi_1 LD_{st} + t_t + \theta_s + \mu'_2 X_{st} + \varepsilon_{st}^1 \quad (4)$$

where  $\theta_s$  are state fixed effects,  $t_t$  denote year fixed effects, and  $X_{st}$  represents additional time-varying state or governor characteristics. We expect policy experimentation to decrease with the incumbent's reputation, i.e. that  $\alpha_1 < 0$ . A politician with high initial reputation will be reluctant to engage in risky policy experiments, while a politician with low reputation is more likely to experiment in the hope of boosting his reelection chances. Our second prediction says that this reputation effect to be weaker for governors who cannot be reelected and hence worry less about their reputation, so  $\pi_1 > 0$ .

Whether governors with low competence are more or less likely to experiment than a governor with high competence is not clear. However, we should observe that governors with low competence are more likely to experiment if they are less concerned with reelection. To test this prediction, we estimate the following model:

$$Experiment_{st} = \alpha_2 LowQuality_{st} + \pi_2 LowQuality_{st} * LD_{st} + \phi_2 LD_{st} + t_t + \theta_s + \mu'_2 X_{st} + \varepsilon_{st}^2 \quad (5)$$

where  $LowQ_{st}$  is equal to one if the governor has little political experience (below the median) and zero otherwise. All other variables are defined as before. We expect that low-quality lame ducks experiment more,  $\pi_2 > 0$ , while the sign of  $\alpha_2$  is theoretically ambiguous.

Our final prediction applies to policy reversals which we test in the following model:

$$Reversal_{st} = \alpha_3 LowQ_{st} + \pi_3 Gains_{st} * LowQ_{st} + \phi_3 * Gains_{st} + t_t + \theta_s + \mu'_3 X_{st} + \varepsilon_{st}^3 \quad (6)$$

where  $Gains_{st}$  measure the potential gains from experimentation and all other variables

are defined as before. Low-quality governors are more likely to revert a policy that failed; hence, we expect that  $\alpha_3 > 0$ . Yet, the higher the gains from a successful experiment, the less likely a policy reversal becomes; so,  $\pi_3 > 0$ . We now discuss our empirical results.

## 6 Empirical Results

### 6.1 Policy Experimentation

Table 1 shows the results for the first prediction. The dependent variable here is the number of waiver applications in the AFDC period (1978-1996). All specifications control for state and year fixed effects as well as governor age to adjust for general experience. In line with our model, we find that governors with high initial reputation among voters experiment less in the waiver period. The relationship becomes slightly stronger when we control for the general contestedness of state elections (e.g. swing states) by including an indicator whether the governor has changed over the past two electoral cycles. Hence, the reputation effect cannot be explained by the fact that swing states with tougher inter-party competition experiment less; rather, experimentation is less likely because the current governor has more to lose if the experiment fails.

Table 1 (columns (3) and (4)) also confirms our second prediction: the reputation effect is indeed much stronger for governors who are not lame ducks. The effect is economically sizable. An increase in the past vote margin by an interquartile range (75p-25p) reduces policy experiments among reelectable governors by 32-35%.<sup>22</sup> For lame ducks in contrast, there is no statistically significant effect of initial reputation on policy experimentation: the sum of the coefficient on the vote margin and the interaction effect of vote margin and lame duck is close to zero for all specifications. The last two columns show that the reputation effect is unchanged if we also control for the potential gain from experimentation. As predicted by the model, higher potential gains (as measured by past welfare spending) are associated with more experimentation. An increase in welfare

---

<sup>22</sup>The interquartile range of the past vote margin is 9.1 percentage points. The mean number of waivers per year in our time period is 0.2. Hence,  $(-0.0076 \cdot 9.1) / 0.2 = -0.3458$ .

spending by an interquartile range increases experimentation by 24%.<sup>23</sup> We find very similar results if we use a binary indicator for a waiver experiment instead (see Table A4 in the appendix).

Do we observe a comparable pattern after the federal welfare reform when states could (and had to) choose their own welfare rules? Table 2 runs the same specification using our measure of policy experimentation for the TANF period. The reputation effect in the sample of all governors is statistically weaker. Yet, once we allow reputation to affect governors who can and cannot be reelected differentially, we find the same pattern as for the waiver period: governors with the option to be reelected in their state experiment less the higher their initial reputation. The economic size of the reputation effect is actually quite similar to the waiver period: an increase in the vote margin by a interquartile range reduces experimentation during the TANF period by about 30-44%.<sup>24</sup> As before, we also find that potential gains from experimentation are associated with more experimentation; and more importantly, controlling for potential gains does not affect the reputation effect for governors. Overall, our findings provide strong evidence that reputational concerns influence the decision to experiment with policy rules before as well as after the US federal welfare reform in 1996.

Table 3 tests our third prediction which links experimentation to the governor's competence. As in Table 1, the dependent variable is the number of waivers a state applied for in a given year over the 1978-1996 period. The first two columns confirm that the competence of the governor is not associated with policy experimentation. The next two columns however, show that governors with low competence who cannot be reelected indeed experiment more. In line with our model, the reputation effect is weaker for governors who cannot be reelected (hence, their relative weight on welfare  $\gamma$  is larger), which makes them more prone to try the risky experiment. The pattern continues to hold and becomes even slightly stronger when we also control for the potential gains

---

<sup>23</sup>The interquartile range of welfare spending during the 1978-1996 period is 237.3. Hence, we get  $237.3 * 0.0003 / 0.2 = 0.237$ .

<sup>24</sup>The interquartile range of the past vote margin is 11.31 percentage points. The mean number of experiments per year is 0.244. Hence,  $(-0.0064 * 11.3) / 0.244 = -0.297$ .

from experimentation.<sup>25</sup> The effects are again economically sizable. Having a lame duck governor with little experience increases policy experimentation by 43-44%.<sup>26</sup>

## 6.2 Policy Reversals

Our final empirical test studies policy reversals. One important caveat is that in the model, the decision to revert a policy experiment is only taken when the experiment does not deliver short-term benefits. Unfortunately, we do not observe whether the experiment failed in the short-term or not. As a consequence, our reversal measure contains substantial measurement error which reduces the precision of our estimates. To reduce this concern, we focus here on substantial changes in the main welfare rules to define a policy reversal.<sup>27</sup>

Table 4 reports estimates from a linear probability model where the dependent variable is whether a state has reversed an experiment in a year or not.<sup>28</sup> Governors with lower competence to recognize whether an experiment is appropriate are more likely to revert an experiment. This result is in line with the model: low-quality governors may implement an experiment in the wrong state of the world and hence, are more likely to revert a policy.

In the second specification, we also include the initial reputation of the governor which, according to our model, should not play a role for the decision to reverse a policy experiment. And indeed, the correlation between initial reputation and a policy reversal is very weak (relative to the effect on policy experimentation in Table 1, for example)

---

<sup>25</sup>The results are statistically weaker, but point in a similar direction for the TANF period (see Table A6). One possible explanation why the results are weaker during the TANF period is that in the model, policy experimentation is always worthwhile ex-ante. This assumption is plausible for the waiver period when states could use the waivers to adjust welfare rules to their local setting. Yet, the assumption might not necessarily be satisfied during the TANF period when states were free to choose their own rules.

<sup>26</sup>Given the binary competence variable, the sum of the main effect and interaction with lame duck is  $(-0.0847+0.1728)/0.2 = 0.441$ .

<sup>27</sup>For work requirements, for example, we code a policy reversal as one if the hours requirement changes by at least 30%. By focusing on substantial policy changes, we hope to reduce measurement error from small policy reversals.

<sup>28</sup>Since the reversal indicator is binary, one could also estimate a probit or logit model. In general, the sign of the coefficients from a probit model look are in line with the ones reported here. Yet, we stick to the linear probability model here as the calculation of marginal effects in a probit model with interaction terms is more involved.

and never statistically significant.

That initial reputation has little relationship with the reversal decision suggests that voters do not seem to update their opinion about a politician based on the observed policy choices (like an experiment). The absence of updating could be an indication that voters may not be able to fully observe or evaluate the politicians' actions and the policy's consequences within a relatively short period of time.<sup>29</sup>

A policy reversal should also be more likely among governors with little experience the lower the potential gains from experimentation. Columns (3) and (4) also find support for this prediction. If the potential gains from experimentation are high, governors with little experience prefer to gamble: they stick to a policy experiment and hope that the experiment turns out to be successful in the second period. Together, these two factors can explain a substantial share of the observed policy reversals.<sup>30</sup>

### 6.3 Robustness Analysis

So far, we have analyzed the decision to experiment and reverse welfare policy rules abstracting from other factors like ideological differences or conflicts between the legislature and the governor. We now discuss a number of validity tests to support our main finding that reputational concerns matter for experimentation in the political arena even conditional on these other influences.

**Ideology.** A common perception of the welfare reform, and certainly the federal PRWORA, is that it was fueled by a conservative agenda of curbing welfare spending.

Casual observation seems to support this: Republicans were typically more in favor of

---

<sup>29</sup>The result that initial reputation does not matter for the decision to revert a policy is different than in the signaling game in Majumdar and Mukand (2004) where voters update their belief about the politician's quality after observing policy choices and their outcomes.

<sup>30</sup>One would also expect that a policy reversal reduces the reelection chances of the governor. Since welfare is negative if a policy experiment is reversed, the reelection probability of a governor who reverts is always zero in our model. Unfortunately, it is quite difficult to test the link between policy reversal and reelection with our data for two reasons. First, we run into small sample issues as we have to exclude lame ducks and governors who do not run in the next election. Second, a policy reversal may only be observed if an experiment failed in the short-term; but failure is not observed by the researcher. In the data, we do find a negative correlation between a policy reversal and reelection for governors who are not lame ducks but the relationship is statistically not very robust due to small sample size and measurement error.

the decision to abolish welfare and encourage workfare. How would ideological differences about the best welfare policy affect our analysis? In the model, voters and politicians agree on which policy is best given the state of the world which is obviously a strong assumption. Our model can be extended to allow for some ideological differences between politicians.<sup>31</sup> One simple way is to assume that Republicans assign a higher probability that welfare reform is the right thing to do, i.e. that the state of the world calls for a reform. Republicans who do not observe the state of the world believe that the state of the world  $N$  (where an experiment is welfare-enhancing) occurs with probability  $p_R > p$ . In contrast, Democrats who do not observe the state of the world continue to have the belief that the state of the world being  $N$  is  $p$ .<sup>32</sup> Since the representative voter cares only about ex-post welfare, the reelection rule remains unaffected by the bias of Republicans. In such environment, it is easy to show that Republicans are more likely to experiment with the reform (which increases expected welfare from Proposition 2) and to persist in the face of short-term failure (which decreases expected welfare from Propositions 3 and 4) relative to a Democratic governor.<sup>33</sup>

Table 5 reruns our baseline analysis from Table 1, but now adds controls for political ideology. Surprisingly, Republican governors are not statistically significantly more likely to experiment than Democratic governors. We find the same result if we use Berry's measure of party ideology (see Berry et al. (1998)) to account for intra-party differences in ideology (column (3)). The absence of an effect of ideology is consistent with other studies which also fail to find an effect of ideology for redistributive politics (see e.g. Aidt and Jensen (2009) for the introduction of the personal income tax). Republican lame ducks are less likely to experiment (see specification (2)) but the coefficient is again not statistically significant. These results provide little support for the idea that lame ducks, in the absence of reelection motives, follow their party ideology; if anything, they seem to go

---

<sup>31</sup>Politicians might also boost their reputation by proposing and implement policies that are on the other side of their ideological spectrum, see Cukierman and Tommasi (1998) and Moen and Riis (2010).

<sup>32</sup>The analysis yields the same prediction if instead Democrats underestimate the probability of the state being  $N$  relative to Republicans.

<sup>33</sup>To illustrate, take the example where  $\gamma = 0$ . Then, from Proposition 3, a politician experiments whenever  $p \geq \lambda$ . Since  $\lambda$  is not correlated with the party of the governor in the model, the inequality is more likely to be satisfied for a Republican. As a consequence, she experiments more.

against their own party. In columns (4) and (5), we test whether policy experimentation is influenced by voter ideology (measured by Berry’s voter ideology measure and the Democratic vote share in the past presidential election) instead. Shifts in voter ideology are apparently not responsible for the decision to experiment. The final specification shows that even including all ideology controls simultaneously does not change our result that reelection concerns matter for policy innovations.

**Legislature.** Our main empirical analysis also restricts attention to the governor as the relevant decision-maker. As we argue above, state governors played a prominent role during the waiver period and the passage of PRWORA. Yet, the legislature is certainly also involved in policy-making. Veto players in the legislative process might delay or reduce policy experimentation while more competition possibly increases it. Previous evidence suggests that party composition and tightness of legislative elections seem to have an effect on welfare spending under AFDC (e.g. Barrilleaux et al. (2002)).

In Table 6, we include a number of controls for the legislature to our baseline specification: the party composition in the lower and upper house; whether the government is divided; the degree of polarization in the parliament (often seen as an indicator for stiff political competition); and finally, the Holbrook and Van Dunk measure of competitiveness. The only statistically significant influence is that divided governments seem to experiment more; all other legislative variables show no relationship with policy experimentation.<sup>34</sup> Across all specifications in Table 6, the coefficient on the past vote margin and its interaction with the lame duck indicator similar in size to the baseline specification in Table 1.

**Spillover Effects.** Our theoretical and empirical analysis also abstracts from spillover effects across states. If potential welfare recipients are geographically mobile, states may engage in welfare competition. In that case, experimentation in one state is influenced by policy choices in neighboring states (see Brueckner (2000), Figlio et al. (1999), Saavedra

---

<sup>34</sup>For a detailed analysis of the effects of divided government on reform-making in US states, see Bernecker (2013).

(2000), or Wheaton (2000)). Spillovers could also arise if policy choices in neighboring states provide information for voters (as in a model of yardstick competition (Besley and Case (1995b)) or politicians. In fact, a central insight of the literature on laboratory federalism is that decentralization promotes learning and spillover of knowledge in a federal system (see Oates (1999) for a survey; Rose-Ackerman (1980); and Strumpf (2002)).

To test for spillover effects, we create three measures of experimentation in neighboring states. Our first measure calculates how much geographic neighbors experiment on average in prior years. Alternatively, we define neighbors as states with similar population size and create their mean propensity to experiment in previous years. Finally, we also consider spillovers from states that are ideologically similar and located within the same census region. Table 7 shows that, once we control for aggregate year effects, there is no systematic statistical relationship between policy experimentation in a state and experimentation in neighboring states over the past two years. Most importantly, cross-state spillovers do not affect our main conclusion that reelection concerns are important determinants of innovation in the public sector.

How does the absence of cross-state spillovers square with the rich case study evidence that certain states and governors (like Tommy Thompson in Wisconsin) played a decisive role in the welfare reform process? The key observation here is that cross-state spillovers matter if we do not control for aggregate waiver activity (through year effects). One possible interpretation is that state governors learn from and possibly imitate neighboring states when there is a lot of experimentation anyways. Yet, once we control for the overall activity pattern of experimentation, spillovers play a much less prominent role.

**Demand for Welfare Services.** Finally, we have not said much about the demand for welfare provisions. One explicit goal of the 1996 welfare reform was to decentralize decision-making to the state level. An important benefit of decentralization is that policies are, under certain conditions, closer to the preferences of the electorate. To test for the influence of the demand side, we add sequentially controls for the unemployment rate,

income per capita, population size and its composition by age and race, the immigrant share and share of unmarried births as well as income inequality (measured by the 90/10 ratio) to our baseline specification. The last specification (in column (5)) adds all socio-demographic controls simultaneously. The set of demographics controls we include are similar to those commonly used in studies of policy innovations (see, for example, Walker (1969), Berry and Berry (1992), or Boehmke and Skinner (2012)). Table 8 shows that larger states experiment more, while states with more immigrants experiment less. We do not find any relationship between the decision to experiment and general economic conditions, population composition or unmarried births.

Overall, our robustness suggests that reelection concerns matter for public sector innovation even after considering ideological considerations, the legislative process, spillover effects and the demand for welfare.

## 7 Conclusion

In this paper, we analyze how reputational concerns among governors influence the decision to experiment and possibly reverse a policy experiment.

Our empirical setting is the US welfare reform in 1996, the most important shift in welfare policy since the New Deal. To guide our empirical analysis, we outline a political agency model and derive several empirical predictions about how reelection concerns affect whether states experiment with welfare rules and possibly reverse them later on. To test our predictions, we build a novel data set of welfare policy experiments and reversals in the US states between 1978 and 2010. We complement our data with detailed information on the governor and legislature as well as additional socio-economic characteristics in each state.

Our predictions confirm that reelection concerns matter for the decision to experiment both during the waiver period before 1996, but also for the period after 1996 when states decide on their welfare policy. In line with our theoretical considerations, governors who face binding time limits (“lame ducks”) behave systematically different from governors

who face reelections and hence worry more about their political reputation. A battery of robustness tests suggests that the link between reelection and experimentation is robust to introducing ideological differences, characteristics of the legislature, spillovers between states, or local demand for welfare services.

## References

- Aidt, T. and P. Jensen (2009). The taxman tools up: An event history study of the introduction of the personal income tax. *Journal of Public Economics* 93, 160–175.
- Alesina, A. and A. Drazen (1991). Why are stabilizations delayed? *American Economic Review* 81(5), 1170–1188.
- Allen, R. and J. Clark (1981). State policy adoption and innovation: Lobbying and education. *State and Local Government Review* 13(1), 18–25.
- Alt, James, E. B. d. M. and S. Rose (2011). Disentangling accountability and competence in elections: Evidence from u.s. term limits. *Journal of Politics* 73(1), 171–186.
- Barrilleaux, C., T. Holbrook, and L. Langer (2002). Electoral competition, legislative balance, and American state welfare policy. *American Political Science Review* 46(2), 415–427.
- Bernecker, A. (2013). Divided we reform? Evidence from US welfare policies. *CESifo Working Paper No. 4564*.
- Berry, F. S. and W. D. Berry (1990). State lottery adoptions as policy innovations: An event history analysis. *American Political Science Review* 84(2), 395–415.
- Berry, F. S. and W. D. Berry (1992). Tax innovation in the states: Capitalizing on political opportunity. *American Journal of Political Science* 36(3), 715–742.
- Berry, F. S. and W. D. Berry (2007). Innovation and diffusion models in policy research. In P. A. Sabatier (Ed.), *Theories of the Policy Process*, second edition, pp. 223–260. Boulder, CO: Westview.
- Berry, W. D., E. J. Ringquist, R. C. Fording, and R. L. Hanson (1998). Measuring citizen and government ideology in the American States, 1960-93. *American Journal of Political Science* 42(1), 327–348.
- Besley, T. and A. Case (1995a). Does electoral accountability affect economic policy choices? Evidence from gubernatorial term limits. *Quarterly Journal of Economics* 110(3), 769–798.
- Besley, T. and A. Case (1995b). Incumbent behavior: Vote-seeking, tax-setting, and yardstick competition. *American Economic Review* 85(1), 25–45.

- Besley, T. and A. Case (2003). Political institutions and policy choices: Evidence from the United States. *Journal of Economic Literature* 41(1), pp. 7–73.
- Bitler, M. and H. Hoynes (2010). The state of the safety net in the post-welfare reform era. *Brookings Papers on Economic Activity Fall*, 71–127.
- Blank, R. M. (2002). Evaluating welfare reform in the United States. *Journal of Economic Literature* 40, 1105–66.
- Boehmke, F. J. and P. Skinner (2012). State policy innovativeness revisited. *State Politics & Policy Quarterly* 12(3), 303–329.
- Brueckner, J. K. (2000). Welfare reform and the race to the bottom: Theory and evidence. *Southern Economic Journal* 66(3), 505 – 525.
- Callander, S. (2011a). Searching and learning by trial and error. *American Economic Review* 101(6), 2277–2308.
- Callander, S. (2011b). Searching for good policies. *American Political Science Review* 105(04), 643–662.
- Callander, S. and B. Harstad (2013). Experimentation in federal systems. *NBER Working Paper No. 19601*.
- Callander, S. and P. Hummel (2014). Preemptive policy experimentation. *Econometrica* 82(4), 1509–1528.
- Center for Economic and Policy Research (2012). *Current Population Survey ORG Uniform Extracts*, Volume Version 1.7. Washington, D.C.
- Ciccone, A. (2004). Resistance to reform: Status quo bias in the presence of individual-specific uncertainty: Comment. *American Economic Review* 94(3), 785–795.
- Cnudde, C. F. and D. J. McCrone (1969). Party competition and welfare policies in the American States. *American Political Science Review* 63, 858–66.
- Coate, S. and S. Morris (1999). Policy persistence. *American Economic Review* 89(5), pp. 1327–1336.
- Council of Economic Advisors (1999). The effects of welfare policy and economic expansion on welfare caseloads: an update. Technical report, Executive Office of the President of the United States.
- Council of State Governments (2012). *Book of the States*. Washington D.C.: Council of State Governments.
- Crouse, G. (1999). Implementation of major changes to welfare policies, 1992–1998. [http://www.aspe.hhs.gov/hsp/Waiver-Policies99/policy\\_CEA.htm](http://www.aspe.hhs.gov/hsp/Waiver-Policies99/policy_CEA.htm) (accessed on February 7, 2014).
- Cukierman, A. and M. Tommasi (1998). When does it take a Nixon to go to China? *American Economic Review* 88(1), 180–197.

- Fang, H. and M. P. Keane (2004). Assessing the impact of Welfare Reform on single mothers. *Brookings Papers on Economic Activity* 1, 1–95.
- Fernandez, R. and D. Rodrik (1991). Resistance to reform: Status quo bias in the presence of individual- specific uncertainty. *American Economic Review* 81(5), 1146–1155.
- Ferraz, C. and F. Finan (2011). Electoral accountability and corruption in local governments: evidence from audit reports. *American Economic Review* 101(4), 1274–1311.
- Figlio, D., V. Kolpin, and W. Reid (1999). Do states play welfare games? *Journal of Urban Economics* 46, 237–254.
- Fu, Q. and M. Li (2014). Reputation-concerned policy makers and institutional status quo bias. *Journal of Public Economics* 110(0), 15 – 25.
- Gagliarducci, S. and T. Nannicini (2013). Do better paid politicians perform better? Disentangling incentives from selection. *Journal of the European Economic Association* 11(2), 369–398.
- Galasso, V. and T. Nannicini (2011). Competing on good politicians. *American Political Science Review* 105(01), 79–99.
- Grogger, J. and L. A. Karoly (2005). *Welfare Reform: Effects of a Decade of Change*. Harvard University Press.
- Harrington, Joseph E., J. (1993). Economic policy, economic performance, and elections. *American Economic Review* 83(1), 27–42.
- Harvey, C., M. J. Camasso, and R. Jagannathan (2000). Evaluating welfare reform waivers under Section 1115. *Journal of Economic Perspectives* 14, 165–188.
- Haskins, R. (2006). *Work over Welfare: The Inside Story of the 1996 Welfare Reform Law*. Brookings Institution Press.
- Holbrook, T. and E. Van Dunk (1993). Electoral competition in the american states. *American Political Science Review* 87(4), 955–962.
- Holmström, B. (1982). Managerial incentive problems - A dynamic perspective. In *Essays in Economics and Management in Honor of Lars Wahlbeck*. Swedish School of Economics.
- Holmström, B. (1999). Managerial incentive problems: A dynamic perspective. *Review of Economic Studies* 66(1), 169–182.
- Jain, S. and S. W. Mukand (2003). Redistributive promises and the adoption of economic reform. *American Economic Review* 93(1), 256–264.
- Kaplan, T. (2000). Wisconsin works. In S. F. Liebschutz (Ed.), *Managing Welfare Reform in Five States: The Challenge of Devolution*. Rockefeller Institute.
- Karch, A. (2007). *Democratic Laboratories: Policy Diffusion among the American States*. University of Michigan Press.

- Klarner, C. (2003). Measurement of partisan balance of state government. *State Politics and Policy Quarterly* 3, 310–319.
- Koerper, K. (1996). Weekly tracking update - welfare reform: Section 1115 Waiver Activity. Office of Planning, Research and Evaluation.
- Kotakorpi, K. and P. Poutvaara (2011). Pay for politicians and candidate selection: An empirical analysis. *Journal of Public Economics* 95(7–8), 877 – 885.
- Leip, D. (2012). Dave Leip’s Atlas of U.S. Presidential Elections. <http://www.uselectionatlas.org>. (accessed on February 7, 2014).
- Lieberman, R. C. and G. M. Shaw (2000). Looking inward, looking outward: The politics of state welfare innovation under devolution. *Political Research Quarterly* 53(2), 215–240.
- List, J. A. and D. M. Sturm (2006). How elections matter: Theory and evidence from environmental policy. *The Quarterly Journal of Economics* 121(4), 1249–1281.
- Majumdar, S. and S. W. Mukand (2004). Policy gambles. *American Economic Review* 94(4), 1207–1222.
- Mead, L. M. (2004). *Government Matters. Welfare Reform in Wisconsin*. Princeton University Press.
- Meyer, B. D. and J. X. Sullivan (2008). Changes in the consumption, income, and well-being of single mother headed families. *The American Economic Review* 98(5), pp. 2221–2241.
- Moen, E. R. and C. Riis (2010). Policy reversal. *American Economic Review* 100(3), 1261–68.
- Moffitt, R. A. (2001). The effect of pre-PRWORA waivers on AFDC caseloads and female earnings, income, and labor force behavior. In S. Danziger (Ed.), *Economic Conditions and Welfare Reform*. W.E. Upjohn Institute.
- Moffitt, R. A. (2002a). Welfare Benefits File. <http://econ.jhu.edu/directory/robert-a-moffitt/>. (accessed on February 7, 2014).
- Moffitt, R. A. (2002b). Welfare programs and labor supply. In A. J. Auerbach and M. Feldstein (Eds.), *Handbook of Public Economics*, Chapter 34, pp. 2393–2430. Elsevier.
- Moffitt, R. A. (2008). Welfare reform: The US experience. *Swedish Economic Policy Review* 14 (2), 11–54.
- Oates (1999). An essay on fiscal federalism. *Journal of Economic Literature* 37(3), 1120–1149.
- Persson, T. and G. E. Tabellini (2000). *Political Economics: Explaining Economic Policy*. Cambridge, Mass.: MIT Press.
- Reintsma, M. (2007). *The Political Economy of Welfare Reform in the United States*. Edward Elgar.

- Rogoff, K. (1990). Equilibrium political budget cycles. *American Economic Review* 80(1), 21–36.
- Rose-Ackerman, S. (1980). Risk taking and reelection: Does federalism promote innovation? *Journal of Legal Studies* 9, 593–616.
- Saavedra, L. (2000). A model of welfare competition with empirical evidence from AFDC. *Journal of Urban Economics* 47, 248–279.
- Smart, M. and D. M. Sturm (2013). Term limits and electoral accountability. *Journal of Public Economics* 107, 93–102.
- Soss, J., S. F. Schram, T. P. Vartanian, and E. O’Brien (2001). Setting the terms of relief: Explaining state policy choices in the devolution revolution. *American Journal of Political Science* 45(2), 378–395.
- Strumpf, K. S. (2002). Does government decentralization increase policy innovation? *Journal of Public Economic Theory* 4(2), 207 – 241.
- U. S. Department of Homeland Security (2011). *Yearbook of Immigration Statistics*. Washington, D.C.: Office of Immigration Statistics.
- United States Census Bureau (2011). Statistical Abstract of the United States: 2012 (131st edition).
- Urban Institute (2012). Welfare Rules Database. Washington, D.C. (accessed on June 6, 2012).
- U.S. Department of Health and Human Services, Administration for Children and Families (1998-2009). Tanf annual reports to congress, i-viii. Technical report, GAO.
- Walker, J. L. (1969). The diffusion of innovations among the American States. *American Political Science Review* 63(3), 880–899.
- Weaver, K. (2000). *Ending Welfare as we Know it*. Brookings Institution.
- Weissert, C. S. (2000). *Learning from Leaders: Welfare Reform Politics and Policy in five Midwestern States*. Rockefeller Institute Press.
- Wheaton, W. C. (2000). Decentralized welfare: Will there be underprovision? *Journal of Urban Economics* 48, 536–555.
- Willems, T. (2013). Political accountability and policy experimentation: Why to elect left-handed politicians? *Mimeo. Princeton University*. mimeo.
- Winston, P. (2002). *Welfare Policymaking in the States: The Devil in Devolution*. Georgetown University Press.

# Appendix

## A Theory

### A.1 Proof of Propositions

**Proof of Proposition 1.** The proof of Proposition 1 follows from the optimization problem of the incumbent politician. A high-quality incumbent politician knows the state of the world. At  $T = 0$ , the incumbent chooses to reform or not (with the option to revert back to  $a_S$  if the reform is undertaken) in order to maximize equation (1).

In state  $S$ , no gain can be obtained from the reform ex-post: the ex-post welfare of the reform in state  $S$  is either  $-c$  or  $-\frac{c}{2}$  if reverted back to  $a_S$ . Therefore, the reform should not be undertaken from a welfare perspective. Since the ex-post welfare of a reform in state  $S$  is negative, the probability of reelection (from the reelection rule) in state  $S$  is (weakly) positive if and only if no reform is undertaken. The optimal decision of the high-quality incumbent is then to stick with the status quo policy  $a_S$  independently of  $\gamma$  and  $\lambda$ .

In state  $N$ , undertaking the reform maximizes ex-post welfare since there are positive welfare gains ex-post with probability 1, i.e. the ex-post welfare with a reform in state  $N$  is equal to  $\Delta - c$  which is strictly positive from (A1), and is never reversed independently of the realization of the benefit in the short-term. In addition, the probability of reelection is also maximized by implementing the reform since a positive ex-post welfare ensures reelection with probability 1, which is (weakly) greater than  $\lambda$  for any  $\lambda \in [0, 1]$ .

**Proof of Proposition 2.** The proof of Proposition 2 follows from the optimization problem of the incumbent politician given by equation (1) for  $\gamma = 1$ . For a low-quality politician who does not know the state of the world, the reform is welfare maximizing from an ex-ante perspective from (A1) and (A2). Therefore, she always implements the reform in  $T = 0$ . However, it is welfare-maximizing to revert back to  $a_S$  if the reform fails in the short-run since the expected future benefit from the reform is negative from (A3).

**Proof of Proposition 3.** The proof of Proposition 3 follows from the optimization problem of the incumbent politician given by equation (1) for  $\gamma = 0$ .

A purely office-motivated politician never reverts back to the status quo after implementing a reform. If the reform delivers benefits in the short-term it is always continued. If the reform does not deliver short-term benefits in  $T = 1$ , a policy reversal back to the status quo implies that the politician would lose the election with probability 1 (because welfare ex-post would be equal to  $-\frac{c}{2}$  and hence, the probability of reelection would be zero from the reelection rule). Whereas if the politician persists with the reform, she can win the election for sure if the long-term benefit realizes which happens with probability  $\beta > 0$ .

For a low-quality politician who does not reverse a reform, the reform maximizes the probability of winning from an ex-ante perspective if it delivers higher chances of reelection than her current reputation. This statement is equivalent to equation (2) where the left-hand side represents the ex-ante winning probability if the reform is undertaken and continued which is equivalent to the probability that the state of the world is  $N$ . The

right-hand side is the ex-ante winning probability if the reform is not undertaken which is equal to the initial reputation of the incumbent.

**Proof of Corollary 1.** The proof of Corollary 1 follows from equation (2).

**Proof of Proposition 4.** The proof of Proposition 4 directly follows from comparing the politician's payoffs in case of reverting the reform and of continuing the reform after a short-term failure. Reverting the reform is at least as good as continuing it if and only if

$$\gamma \left( \frac{-c}{2} \right) \geq \gamma (\beta \Delta - c) + (1 - \gamma) \beta,$$

or

$$-\beta \geq \gamma \left( \beta \Delta - \frac{c}{2} - \beta \right).$$

Since the term in parenthesis is negative from (A3), the last inequality can be rewritten to obtain the threshold  $\gamma^*$  in equation (3).

**Proof Corollary 2.** The proof of Corollary 2 directly follows from differentiating equation (3) with respect to  $\gamma$  and  $\Delta$ , respectively.

## A.2 Intermediate politician and decision to experiment

Here, we analyze the case whether a low-quality incumbent who cares about both welfare and reelection ( $\gamma \in [0, 1]$ ) undertakes the reform.

**Proposition 5.** (I.) *If  $\gamma \geq \gamma^*$ , then an incumbent politician with low-quality always undertakes a reform if and only if*

$$\gamma \left( pq\Delta - \frac{c}{2}(1 + pq) \right) + (1 - \gamma)pq \geq (1 - \gamma)\lambda. \quad (7)$$

(II.) *If  $\gamma < \gamma^*$ , then an incumbent politician with low-quality always undertakes a reform if and only if*

$$\gamma(p\Delta - c) + (1 - \gamma)p \geq (1 - \gamma)\lambda. \quad (8)$$

*Proof.* The proof of Proposition 5 directly follows from the comparison of the politician's payoff if she undertakes the reform (for the two cases where she will or will not revert back to the status quo in the face of short-term failure) with the payoff from sticking with the status quo and running the next election on her initial reputation instead.  $\square$

Proposition 5 shows that a low-quality incumbent compares the possible welfare and reelection gains that experimentation will deliver (taking into account her optimal reversal decision) to her current chances of reelection determined by her initial reputation.

**Corollary 3.** *A low-quality incumbent is more likely to experiment when  $\Delta$  increases,  $\gamma$  increases and  $\lambda$  decreases.*

*Proof.* The proof of Corollary 3 follows directly from differentiating equations (7) and (8) with respect to  $\Delta$ ,  $\gamma$  and  $\lambda$  respectively.  $\square$

## B Data

### B.1 Welfare Policy Rules during the TANF period

To measure policy experimentation and reversals for the post-1996 period, we rely on the Welfare Rules Database by the Urban Institute. The Welfare Rules Database (WRD) was developed to provide detailed information about states' TANF policies obtained from caseworker manuals and regulations. The database is often more detailed and up-to-date than the official plans of state choices under the block grant that states submit periodically to the federal government. The database contains literally hundreds of rules on eligibility, benefit calculation, and many other aspects of welfare reform. We focus on a set of rules in policy areas that were at the center of the public and political debate surrounding welfare reform. Table A2 provides an overview of the rules and how we coded policy experiments and policy reversals.

**Family caps.** Under AFDC, benefit levels increased with family size. Hence, when a child was born to a member of an assistance unit, the benefit increased to meet the needs of the new child. Under a family cap, additional benefits an assistance unit would receive for adding another member to the unit would be capped. Some states provide a percentage of the increase to the unit, while others provide no additional funds to the unit for the additional child. In several states, a family is never able to regain benefits for a capped child, even after the case has been closed for a period of time. In others, a family cap can be removed (and hence, the child can be included in the benefit computation should the family apply for assistance again) if the assistance unit remains off welfare for some time. We code a binary indicator equal to one if a state has a family cap and zero if not.

**Work requirements.** Under AFDC, states could require recipients to participate in the Job Opportunities and Basic Skills Training (JOBS) program, which provided education, training, and work experience activities. However, many individuals were exempt from these requirements (because of age, illness or having a small child). Under TANF, states require adults heading an assistance unit to perform some type of work-related activity. Work programs vary widely from state to state in terms of who must work, how much work is required, and what activities are considered work. The first rule defines the minimum number of hours a recipient must participate in work-related activities. The hours requirements vary from a mere effort to find a job up to full-time employment. The second rule defines whether the work requirement applies after several months of benefit receipt or by the time of application or approval. The third rule indicates whether there is a time limit of benefit receipt if a parent fails to work at least 20 hours per week in a regular job after a certain number of assistance months. The fourth rule counts the number of work exemptions due to, for example, pregnancy, disability or caring for a young child or elderly person.

**Sanctions.** Under AFDC, sanctions were sometimes imposed if the adult heading an assistance unit did not comply with the JOBS program; in that case, the adult portion of the benefit was not paid out to the assistance unit. Under TANF, states now require household heads to perform some type of work-related activity sooner or later. If a benefit unit does not comply with these requirements, states can impose drastic sanctions. The

first policy rule characterizes the initial sanction if a benefit unit fails to comply with the work requirements for the first time. The initial sanction varies from reduction of 25% or less to a suspension of the full family benefit. The other three rules characterize the severity of the worst sanction that can be imposed. The second rule defines the severity of the worst sanction varying from less than 25% of the benefit to a suspension of the full family benefit and even case closure. The third rule defines the duration of the worst sanction ranging from until the unit complies with the requirement to a permanent suspension of the family benefit. The final rule defines whether a unit has to reapply (or not) after the worst sanction has been imposed.

**Termination and work-requirement time limits.** Under AFDC, families were entitled to receive benefits as long as they met the eligibility requirements. Under TANF, many states imposed both intermittent and lifetime time limits. The first rule characterizes the number of months an assistance unit can receive benefits without interruption (many states impose 24 months). The second rule defines how benefits are reduced when the assistance unit reaches the intermittent time limit. The loss of benefits might be just for the adult members or for the entire assistance unit. The third rule defines the duration of the lifetime limit ranging from no time limit to only 24 months. The federal government has itself imposed a 60 months lifetime limit; states that wish to extend benefit receipt beyond five years have to use their own state funds to finance it. The final rule defines whether the state allows for any extensions to the lifetime limit or not.

**Earnings disregards (extended set of rules).** Under AFDC, all recipients who worked were entitled to a \$90 work expense disregard. In addition, for the first four months of AFDC receipt, the next \$30 of earned income, plus one-third of the remainder, was disregarded in calculating eligibility and benefits. After four months and until one year, only the \$30 disregard continued. After one year, there was no earned income disregard. Hence, after one year of AFDC receipt, the benefit amount was reduced by one dollar for every dollar that a beneficiary earned above the amount set aside to cover her work expenses. Many states relaxed those rules during the TANF period. We calculate the earnings disregards a state allows in month 5 for an individual who works full-time earning the minimum wage.

**Mandatory job search and diversion payments (extended set of rules).** Under AFDC, applicants were not required to engage in job search activities in order to be eligible for benefits. After 1996, certain states require applicants to engage in job search even before applying to TANF as a necessary precondition for eligibility. Since 1996, many states also offer a one-time cash payment in lieu of a TANF monthly benefit to assist families or individuals with short run income fluctuations. If individuals decide to take up this payment, they cannot apply for TANF for a certain period of time. Diversion payments did not exist during the AFDC period. We construct two binary indicators equal to one if the state requires mandatory job search or offers diversion payments to keep people off welfare.

## B.2 Politics and ideology measures

**Governor characteristics.** Information on US governors comes from the website of the National Governor’s Association merged with data kindly provided by David J. Andersen from the Eagleton Institute of Politics at Rutgers University. We calculate a governor’s competence from her political experience prior to becoming governor. Specifically, the variable measures the number of years between the first electoral office (such as member of the State Senate or Attorney General) and the inauguration as governor. This gives a continuous quality measure. We then construct a low quality dummy equal to one for governors whose prior experience is being below the median and zero otherwise. We also have information on the age of the governor from the same sources. Information on gubernatorial election results (especially vote margins) and term limits are taken from List and Sturm (2006); Council of State Governments (2012); and Leip (2012)).

**State Legislature.** Data for the composition of the state legislature, state competitiveness, the party of the governor and indicators for a divided government are obtained from Klarner (2003) and updated using Klarner’s webpage (<http://www.indstate.edu/polsci/klarnerpolitics.htm>). Polarization in the state legislature is calculated as  $|\text{democratic seat share} - 50\%|$  for the state senate and house respectively. The divided government indicator is equal to one if the governor belongs to a different party than the majority of legislators in either the state senate or the state house. To measure state competitiveness, we use the Holbrook-Van Dunk index calculated from district-level returns to state legislative elections (see Holbrook and Van Dunk (1993) for details).

**Ideology.** The vote share for the Democratic candidate in the last presidential election is taken from the Statistical Abstract (United States Census Bureau (2011)) and updated using Leip (2012). We also use voter and government ideology based on ideology ratings of the state’s congressional delegation, the American for Democratic Action (ADA) rating and the AFL/CIO’s Committee on Political Education (COPE) rating (Berry et al. (1998)). Berry et al. (1998) assign an ideology rating to the citizens of each congressional district using a weighted average of the score of the congressional member and his or her election opponent, weighting the scores according to the number of votes they received. Zero denotes the most conservative and 100 the most liberal. They then generate a state-wide measure by averaging over all congressional districts. The measure of government ideology is constructed by assigning to the governor and major party delegations in the legislature the ratings of the members of Congress from their party. Updates of these ideology data are available at <http://rcfording.wordpress.com/state-ideology-data/>.

## B.3 State Demographics and Other Controls

**Demographics.** Population size, the number of Blacks and the age structure are taken from the Statistical Abstract (United States Census Bureau (2011)). The size of the immigrant population refers to the number of legal immigrants admitted by state of intended residence is taken from Fang and Keane (2004) for 1970 to 2002 and updated using the Yearbook of Immigration Statistics (U. S. Department of Homeland Security (2011)) which is available online at <http://www.dhs.gov/yearbook-immigration-statistics>. Personal income per capita and wage income is taken from the website of the Bureau of

Economic Analysis. The unemployment rate for 1960 to 1998 is from Moffitt (2002a) and updated for 1998 to 2010 from the website of the Bureau of Labor Statistics. Our measure of income inequality is the ratio of the 90th over the 10th percentile of total household income calculated from the March Current Population Survey (Center for Economic and Policy Research (2012)). All income variables are deflated by the urban consumer price index with base year 2002. Unmarried Birth refers to the % of all births to unmarried women per 1,000 unmarried women aged 15-44 years by state of residence. For the years 1992 to 2003, the data are available from Table 8.3 in the TANF Annual Reports to Congress (U.S. Department of Health and Human Services, Administration for Children and Families (2009)). For earlier and later years, we obtain the data from the Center for Disease Control and Prevention (CDC), the National Vital Statistics System and Vital Statistics available from the National Bureau of Economic Research.

**State welfare caseloads and finances.** TANF caseloads for both fiscal year and calendar year are taken from Moffitt (2002a) for 1960 to 1998 and updated to 2010 using the Statistical Abstract (United States Census Bureau (2011)). To measure the potential gains from experimentation, we use state welfare spending. These data come from Paul Ehmann at the US Census Bureau. All fiscal variables are then converted into real measures using the urban consumer price index (with years 1982-1984=100) from the Bureau of Labor Statistics.

**Spillover effects across States.** To analyze spillover effects across states, we define a “neighbor” along three different dimensions. The first variable (geographic neighbors) codes the extent of experimentation for all states sharing a common border with the current state. The second variable (similar population sizes) calculates experimentation in states with a similar population size as the current state. For the calculation, we use the following ten bands based on population size in 1978: (CA NY TX PA IL), (OH MI FL NJ MA), (NC IN GA VA MO), (WI TN MD LA MN), (WA AL KY CT SC), (IA OK CO AZ OR), (MS KS AR WV NE), (UT NM ME RI HI), (ID NH MT NV SD) and (ND DE VT WY AK). The third measure (ideological neighbors) takes the average number of experiments of the four states within the same census region who voted similarly in the presidential election of 1978.

**Table 1: Initial Reputation and Policy Experimentation during the Waiver Period**

	<u>Main Effect</u> <u>Reputation</u>		<u>Differential Effect</u> <u>for Lame Ducks</u>		<u>Control for</u> <u>Potential Gains</u>	
	(1)	(2)	(3)	(4)	(5)	(6)
Past Governor Vote Margin	-0.0046*	-0.0052*	-0.0074**	-0.0076**	-0.0066*	-0.0070*
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
Lame Duck * Past Governor Vote Margin			0.0082*	0.0082	0.0083	0.0088
			(0.005)	(0.005)	(0.007)	(0.007)
Lame Duck			-0.0236	-0.0364	-0.0371	-0.0496
			(0.065)	(0.073)	(0.088)	(0.091)
Past AFDC Spending					0.0003***	0.0002***
					(0.000)	(0.000)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Governor Age	Yes	Yes	Yes	Yes	Yes	Yes
Change in Governor in Past Elections	No	Yes	No	Yes	No	Yes
Observations	872	827	872	827	785	785
R-squared	0.278	0.278	0.281	0.280	0.292	0.293

*Notes* : The dependent variable in all specifications is the number of policy experiments (waiver applications) in a given state and year. Columns (1) and (2) include the governor's vote margin in the last gubernatorial election (measured as the distance to the runner-up in percent). Columns (3) and (4) allow the effect of the vote margin to differ for governors who cannot be reelected ("lame ducks"). Columns (5) and (6) also control for AFDC spending (measured in US\$ million) two years earlier as a proxy for potential gains from experimentation. All specifications control for year and state fixed effects as well as governor age. Even columns add an indicator whether the governor has changed over the past two electoral cycles. Standard errors clustered at the state level are shown in parentheses. \*\*\* p<0.01, \*\* p<0.05 and \*p<0.1.

**Table 2: Initial Reputation and Policy Experimentation during the TANF Period**

	<u>Main Effect</u> <u>Reputation</u>		<u>Differential Effect for</u> <u>Lame Ducks</u>		<u>Control for</u> <u>Potential Gains</u>	
	(1)	(2)	(3)	(4)	(5)	(6)
Past Governor Vote Margin	-0.0036 (0.002)	-0.0036 (0.002)	-0.0064** (0.003)	-0.0064** (0.003)	-0.0077*** (0.003)	-0.0078*** (0.003)
Lame Duck * Past Governor Vote Margin			0.0052 (0.005)	0.0052 (0.005)	0.0055 (0.005)	0.0055 (0.005)
Lame Duck			-0.0125 (0.098)	-0.0121 (0.099)	-0.0382 (0.110)	-0.0374 (0.109)
Past TANF Spending					0.0003* (0.000)	0.0003* (0.000)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Governor Age	Yes	Yes	Yes	Yes	Yes	Yes
Change in Governor in Past Elections	No	Yes	No	Yes	No	Yes
Observations	687	687	687	687	544	544
R-squared	0.413	0.413	0.413	0.413	0.339	0.339

*Notes* : The dependent variable in all specifications is the number of policy experiments (changes in welfare rules) in a given state and year. Columns (1) and (2) include the governor's vote margin in the last gubernatorial election (measured as the distance to the runner-up in percent). Columns (3) and (4) allow the effect of the vote margin to differ for governors who cannot be reelected ("lame ducks"). Columns (5) and (6) also control for TANF spending (measured in US\$ million) two years earlier as a proxy for potential gains from experimentation. All specifications control for year and state fixed effects as well as governor age. Even columns add an indicator whether the governor has changed over the past two electoral cycles. Standard errors clustered at the state level are shown in parentheses. \*\*\* p<0.01, \*\* p<0.05 and \*p<0.1.

**Table 3: Governor Quality and Policy Experimentation during the Waiver Period**

	<u>Main Effect</u>		<u>Differential Effect</u>		<u>Control for</u>	
	<u>Governor Competence</u>		<u>for Lame Ducks</u>		<u>Potential Gains</u>	
	(1)	(2)	(3)	(4)	(5)	(6)
Governor with Low Competence	-0.0380 (0.064)	-0.0378 (0.063)	-0.0851 (0.075)	-0.0847 (0.074)	-0.0999 (0.081)	-0.0972 (0.080)
Lame Duck * Governor with Low Competence			0.1782** (0.080)	0.1728** (0.079)	0.2014** (0.094)	0.1993** (0.095)
Lame Duck			-0.0259 (0.049)	-0.0081 (0.050)	-0.0449 (0.061)	-0.0334 (0.062)
Past Governor Vote Margin		-0.0045 (0.003)		-0.0049* (0.003)		-0.0046 (0.003)
Past AFDC Spending					0.0003*** (0.000)	0.0003*** (0.000)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Governor Age	Yes	Yes	Yes	Yes	Yes	Yes
Observations	872	872	872	872	785	785
R-squared	0.276	0.278	0.280	0.283	0.293	0.295

*Notes:* The dependent variable in all specifications is the number of policy experiments (waiver applications) in a given state and year. Columns (1) and (2) includes an indicator whether a governor is below the median political experience in the governor sample (as a proxy for the governor's competence). Columns (3) and (4) allow the effect of governor competence to differ for governors who cannot be reelected ("lame ducks"). Columns (5) and (6) also control for AFDC spending (measured in US\$ million) two years earlier as a proxy for potential gains from experimentation. All specifications control for year and state fixed effects as well as governor age. Even columns add the governor's vote margin in the last gubernatorial election (measured as the distance to the runner-up in percent). Standard errors clustered at the state level are shown in parentheses. \*\*\* p<0.01, \*\* p<0.05 and \*p<0.1.

**Table 4: Electoral Incentives and Policy Reversals in the TANF Period**

	<u>Main Effect</u>		<u>Effect for</u>	
	<u>Governor Competence</u>		<u>Potential Gains</u>	
	(1)	(2)	(3)	(4)
Governor with Low Competence	0.0531**	0.0520**	0.0618***	0.0593**
	(0.021)	(0.022)	(0.023)	(0.024)
Governor Low Competence * Past TANF Spending			-0.0014***	-0.0013**
			(0.0052)	(0.0053)
Past TANF Spending (/100)			0.0022	0.0021
			(0.003)	(0.004)
Past Governor Vote Margin (/100)		0.0011		0.0010
		(0.001)		(0.001)
Year Fixed Effects	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes
Governor Age	Yes	Yes	Yes	Yes
Observations	697	676	697	676
R-squared	0.112	0.115	0.113	0.116

*Notes:* The dependent variable in all specifications is an indicator whether the state has reversed a policy rule in a given year or not. Columns (1) and (2) include an indicator whether a governor has below median political experience in the governor sample (as a proxy of the governor's competence). Columns (3) and (4) include past state TANF spending (measured in US\$ million) as a measure of the potential gains from experimentation. The coefficients on past TANF spending are divided by 100. All specifications control for state and year fixed effects as well as governor age. Standard errors clustered at the state level are shown in parentheses. \*\*\* p<0.01, \*\* p<0.05 and \*p<0.1.

**Table 5: The Role of Ideology for Policy Experimentation**

	(1)	(2)	(3)	(4)	(5)	(6)
Governor's Past Vote Margin	-0.0075** (0.004)	-0.0069* (0.004)	-0.0074** (0.004)	-0.0074** (0.004)	-0.0076** (0.004)	-0.0081** (0.004)
Lame Duck * Past Vote Margin	0.0082* (0.005)	0.0079 (0.006)	0.0082* (0.005)	0.0082* (0.005)	0.0082* (0.005)	0.0086* (0.005)
Lame Duck	-0.0237 (0.064)	-0.0054 (0.096)	-0.0242 (0.065)	-0.0234 (0.064)	-0.0237 (0.064)	-0.0185 (0.080)
Republican Governor	0.0032 (0.034)	-0.0001 (0.040)			-0.0749 (0.134)	-0.0703 (0.157)
Lameduck * Republican Governor		-0.0604 (0.072)			-0.0424 (0.064)	-0.0532 (0.076)
Governor Party Ideology (Berry et al. 1998)			0.0171 (0.066)			0.1466 (0.247)
Citizens Ideology (Berry et al. 1998)				-0.0003 (0.003)		0.0002 (0.003)
Democratic Presidential Vote Share					0.0024 (0.005)	0.0005 (0.005)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Governor Age	Yes	Yes	Yes	Yes	Yes	Yes
Observations	872	774	872	872	872	872
R-squared	0.281	0.292	0.281	0.281	0.281	0.282

*Notes:* The dependent variable in all specifications is the number of policy experiments (waiver period) in a given state and year. The past vote margin is measured for the last gubernatorial election in percent. The ideology measure for the government is calculated from Berry et al. 1998 and ranges from zero for most conservative to 100 for most liberal. The citizen ideology measure is also taken from Berry et al. 1998; see the data appendix for further details. Democratic presidential vote share refers to the last presidential election. All specifications control for year and state fixed effects as well as governor age. Standard errors clustered at the state level are shown in parentheses. \*\*\* p<0.01, \*\* p<0.05 and \*p<0.1.

**Table 6: Other Political Factors and Policy Experimentation**

	(1)	(2)	(3)	(4)	(5)
Governor's Past Vote Margin	-0.0073* (0.004)	-0.0069* (0.004)	-0.0065* (0.004)	-0.0067* (0.004)	-0.0062 (0.004)
Lame Duck * Past Vote Margin	0.0084* (0.005)	0.0085* (0.005)	0.0088* (0.005)	0.0085* (0.005)	0.0087* (0.005)
Lame Duck	-0.0191 (0.065)	-0.0278 (0.065)	-0.0246 (0.065)	-0.0258 (0.065)	-0.0228 (0.067)
Dem. Seat Share Upper House	-0.3082 (0.291)				-0.1225 (0.293)
Dem. Seat Share Lower House	-0.0131 (0.320)				0.0570 (0.321)
Divided Government		0.0734** (0.036)			0.0479 (0.037)
Polarization Upper House			-0.3970 (0.350)		-0.3638 (0.373)
Polarization Lower House			-0.2910 (0.363)		-0.1962 (0.353)
Competitiveness Index				0.0045 (0.004)	-0.0005 (0.004)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes
Governor Age	Yes	Yes	Yes	Yes	Yes
Observations	872	864	872	861	853
R-squared	0.282	0.283	0.284	0.283	0.287

*Notes* : The dependent variable in all specifications is the number of policy experiments (waiver period) in a given state and year. The past vote margin is measured for the last gubernatorial election in percent. Divided Government is equal to one if the party of the governor is different from the party of the majority of legislators in either the state's lower or upper house. The polarization variables are calculated as absolute deviations of the democratic seat share from 50%. Competitiveness (based on Holbrook and Van Dunk) is measured as the average percent than winning candidates in state legislative elections obtained over the current and past 3 years. All specifications control for state and year fixed effects as well as governor age. Standard errors clustered at the state level are shown in parentheses. \*\*\* p<0.01, \*\* p<0.05 and \*p<0.1.

**Table 7: Spillovers between States and Policy Experimentation**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Governor's Past Vote Margin	-0.0071*	-0.0068*	-0.0071**	-0.0068*	-0.0072*	-0.0067*	-0.0067*
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Lame Duck * Past Vote Margin	0.0076	0.0076	0.0077	0.0076	0.0080	0.0075	0.0073
	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)	(0.006)
Lame Duck	-0.0194	-0.0150	-0.0247	-0.0161	-0.0256	-0.0145	-0.0139
	(0.068)	(0.078)	(0.071)	(0.078)	(0.069)	(0.078)	(0.081)
Geographic Neighbors' Experiments (t-1)	-0.1117						-0.1475*
	(0.082)						(0.088)
Geographic Neighbors' Experiments (t-2)		-0.0714					-0.1020
		(0.092)					(0.093)
Pop. Size Neighbors' Experiments (t-1)			0.2091				0.2154
			(0.188)				(0.190)
Pop. Size Neighbors' Experiments (t-2)				0.0859			0.0930
				(0.165)			(0.156)
Ideological Neighbors' Experiments (t-1)					-0.0487		-0.0003
					(0.098)		(0.087)
Ideological Neighbors' Experiments (t-2)						0.0139	0.0109
						(0.095)	(0.068)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Governor Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	827	785	827	785	827	785	785
R-squared	0.281	0.285	0.286	0.286	0.279	0.285	0.296

*Notes:* The dependent variable in all specifications is the number of policy experiments (waiver period) in a given state and year. The past vote margin is measured for the last gubernatorial election in percent. Experiments by geographic neighbors refers to the average value of the dependent variable for adjacent states (sharing a boundary). Experiments by population size neighbors refers to the average number of waivers in states with similar population size (where states are divided into ten groups of similar population size). Experiments by ideological neighbors refers to the average number of waivers for the four states within the same census region which voted most similarly in the 1978 presidential election. All specifications control for year and state fixed effects as well as governor age. Standard errors clustered at the state level are shown in parentheses. \*\*\* p<0.01, \*\* p<0.05 and \*p<0.1.

**Table 8: Socioeconomic Characteristics and Policy Experimentation**

	(1)	(2)	(3)	(4)	(5)
Governor's Past Vote Margin	-0.0069*	-0.0073**	-0.0064*	-0.0075**	-0.0055
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Lame Duck * Past Vote Margin	0.0084*	0.0086*	0.0075	0.0081*	0.0083
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
Lame Duck	-0.0322	-0.0256	-0.0283	0.0000	0.0000
	(0.070)	(0.065)	(0.074)	(0.000)	(0.000)
				-0.0000***	-0.0000**
Unemployment Rate	-0.0088			(0.000)	(0.000)
	(0.013)				(0.014)
Per Capita Income (in \$1,000)	-0.0000				-0.0000
	(0.000)				(0.000)
Population (in 1,000)	0.0870***				0.1255***
	(0.027)				(0.042)
% Population Black		-0.0231			0.0215
		(0.059)			(0.060)
% Population 65+		-0.0570			-0.0216
		(0.059)			(0.065)
% Immigrant Population			-0.0408***		-0.0569***
			(0.006)		(0.011)
% Unmarried Births			-0.0101		-0.0081
			(0.012)		(0.012)
Income Inequality				-0.0255	-0.0291
				(0.019)	(0.021)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes
Governor Age	Yes	Yes	Yes	Yes	Yes
Observations	872	872	817	872	817
R-squared	0.290	0.282	0.290	0.282	0.299

*Notes* : The dependent variable in all specifications is the number of policy experiments (waiver period) in a given state and year. The past vote margin is measured for the last gubernatorial election in percent. Income inequality is measured as the 90th percentile over the 10th percentile within each state and year. All specifications control for state and year fixed effects as well as governor age. Standard errors clustered at the state level are shown in parentheses. \*\*\* p<0.01, \*\* p<0.05 and \*p<0.1.

Figure 1: AFDC Caseload per Capita 1970-2000

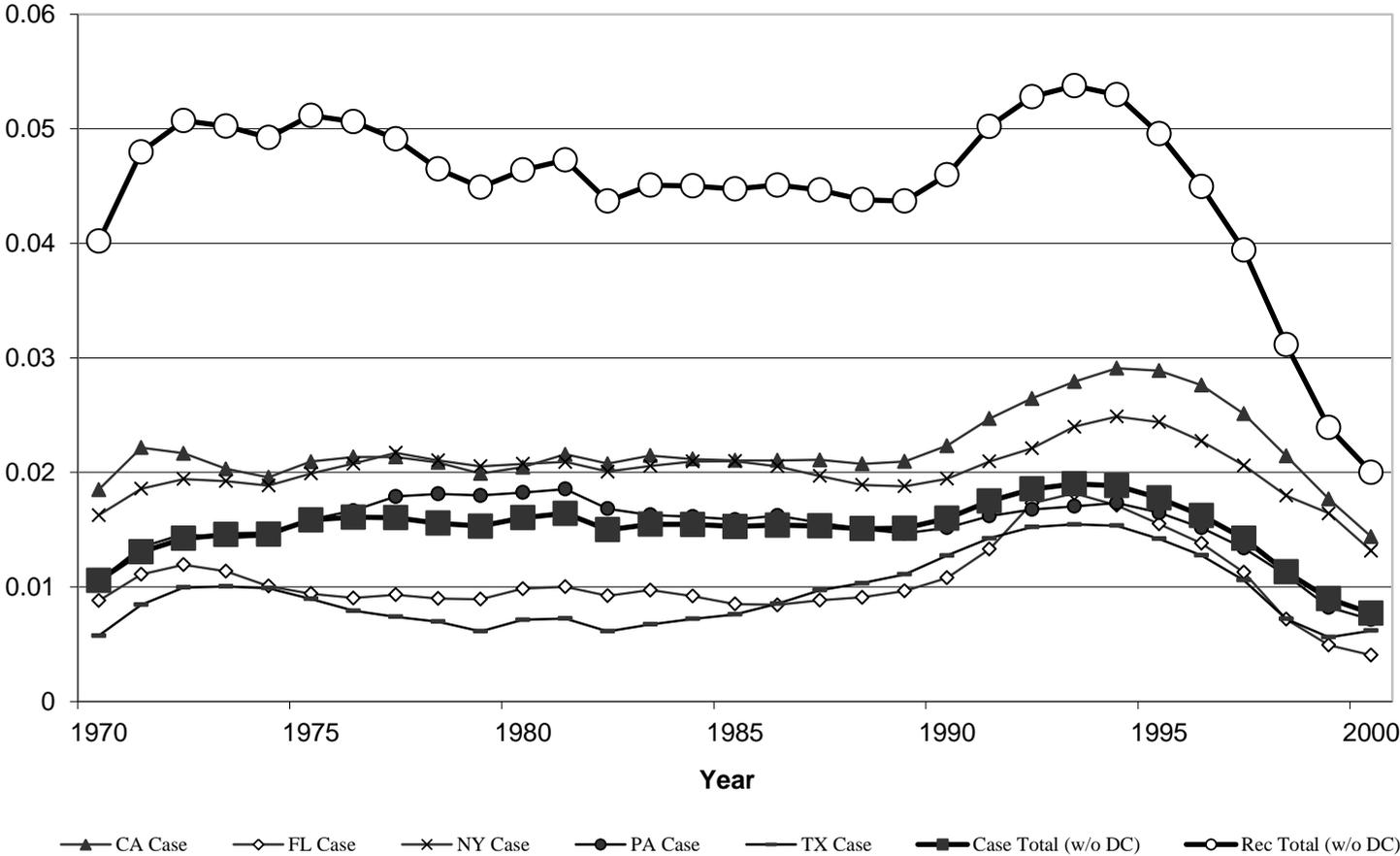


Figure 2: Evolution of Welfare Waivers (1978-1996)

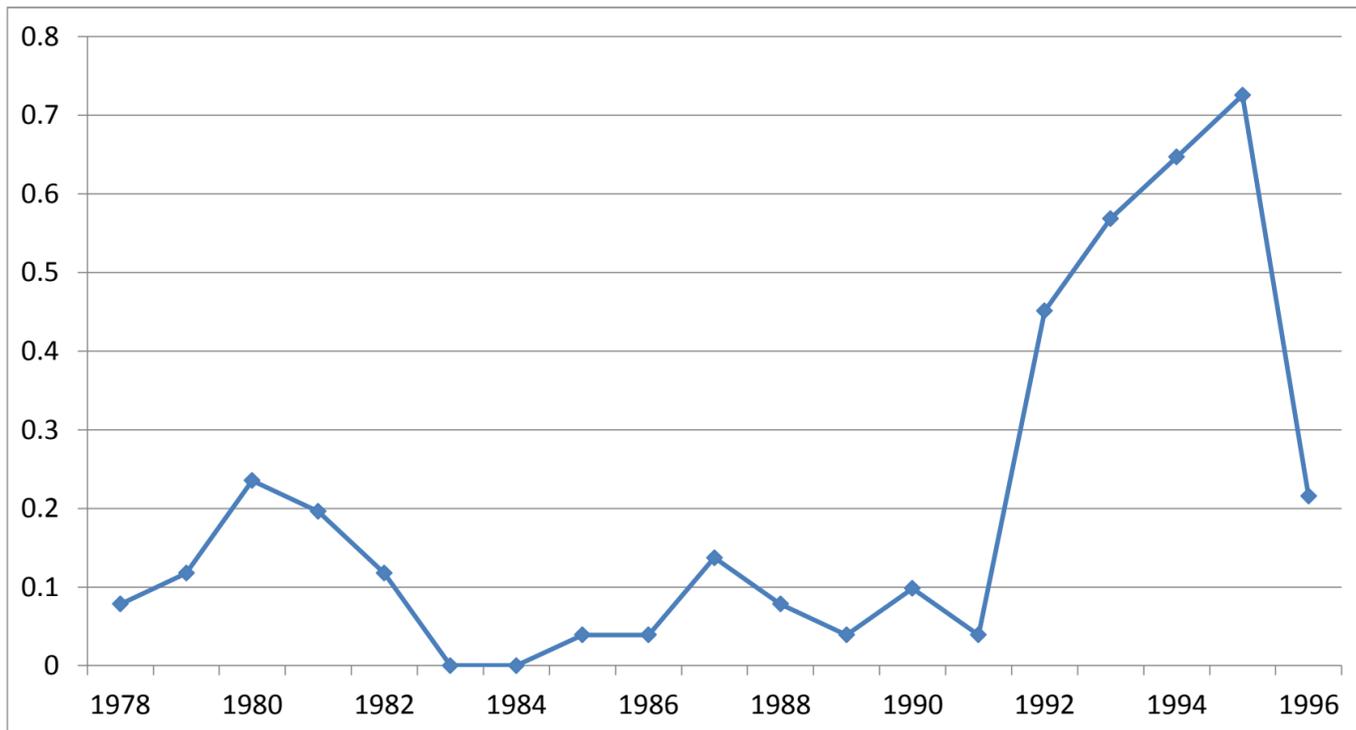
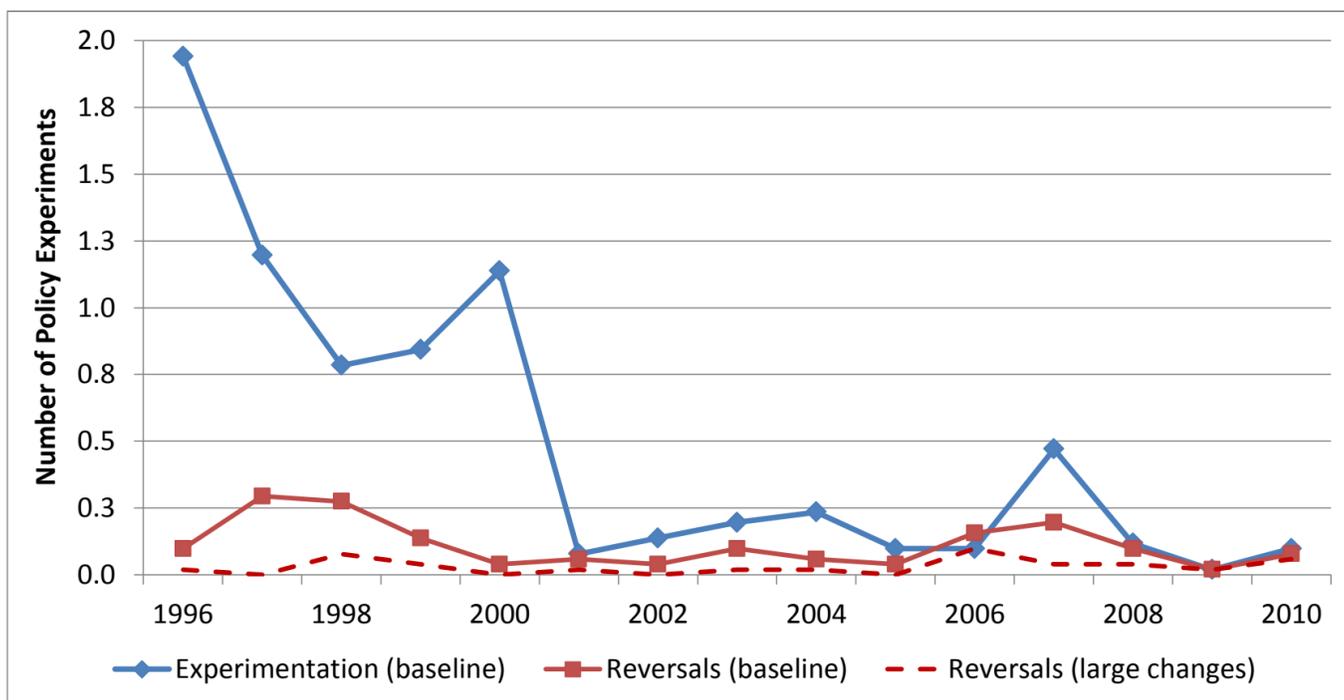


Figure 3: Evolution of Policy Experiments and Reversals (1996-2010)



Source : Authors' calculation; see the data appendix and Table A2 for a description of the welfare rules.

**Table A1: Temporal and Geographic Distribution of Welfare Waivers (1978-1996)**

Year	# States with Waiver Application	State Codes of Waiver Applicants	Average # of Waivers States Applied for
1978	4	CA, UT, VT, WI	1.0
1979	5	Co, MA, MI, NY, WI	1.2
1980	8	CA, CO, CT, FL, GA, IL, NJ, WV	1.5
1981	9	FL, IL, KY, MA, NJ, OH, TX, WI, WV	1.1
1982	6	NY, OH, OK, TX, WV, WY	1.0
1983	0	-	0.0
1984	0	-	0.0
1985	2	CA, IL	1.0
1986	2	CA, MN	1.0
1987	4	IA, MS, WA, WI	1.8
1988	4	AL, CA, NY, OH	1.0
1989	2	CA, MD	1.0
1990	5	GA, MN, OK, UT, WI	1.0
1991	2	OH, VA	1.0
1992	15	CA, GA, IL, MD, MI, MO, NJ, OK, OR, SC, UT, VA, VT, WI, WY	1.5
1993	23	AR, CA, CO, CT, FL, GA, HI, IA, IL, MA, MO, MS, ND, NH, NM, OH, OR, SD, TX, VA, WA, WI, WY	1.3
1994	23	AR, AZ, CA, GA, IL, IN, KS, MA, MD, ME, MI, MO, MT, ND, NM, NY, OH, OK, OR, PA, SC, VA, WI	1.4
1995	24	CT, DE, FL, GA, HI, IL, IN, LA, MA, ME, MO, MS, NC, NH, OH, OK, OR, SC, TX, UT, VA, WA, WI, WV	1.5
1996	7	CA, IA, MD, MI, MN, TN, UT	1.6

Sources : Lieberman and Shaw (2000), Koerper (1996), Crouse (1999).

**Table A2: Coding of Policy Experimentation and Reversals during the TANF Period (1996-2010)**

Policy Rule	Description of Rule	Experimentation = 1	Reversal = 1
<b><u>Family Cap (1)</u></b>	Benefits do not increase if an additional child is born in family while receiving benefits.	State adopts family cap before or after 1996.	State abolishes family cap in 1996 or later.
<b><u>Work Requirements (3)</u></b>			
Hours Requirement	Minimum # of hours a recipient must participate in work-related activities.	Increase in minimum number of working hours required of the recipient.	State reduces or abolishes minimum number of working hours required of the recipient.
Work upon Enrollment	Work requirements apply at application, approval or upon benefit receipt (or only later).	State requires work requirement to be fulfilled at application, approval or upon benefit receipt.	State allows work requirement to be fulfilled later.
Time Limit to Work	Work at least 20 hours per week in an unsubsidized job after a certain period of benefit receipt.	State adopts a time limit to fulfill the requirement of working in a unsubsidized job for at least 20 hours.	State abolishes time limit to fulfill the requirement of working in an unsubsidized job for at least 20 hours.
<b><u>Time Limits (3)</u></b>			
Duration of Lifetime Limit	Maximum # of months an assistance unit can receive benefits over the lifetime.	Prior to 1997, if a time limit is adopted; after 1997, when a state adopts a time limit stricter than the 60 months federal time limit. In 1997, if a state adopts a time limit other than the federal time limit.	If a state abolishes the time limit or increases # of months.
Intermittent Time Limit	# months time an assistance units can receive benefits without interruption.	State reduces intermittent time limit.	State increases intermittent time limit.
Benefit Reduction after Intermittent Time Limit	How much benefits are reduced (adult portion or benefit unit) when assistance units hits the	State reduces benefits more after unit hits intermittent time limit.	State reduces benefits less after unit hits intermittent time limit.
<b><u>Sanctions (4)</u></b>			
Severity of Worst Sanction	How much benefits are reduced under worst sanction for non-compliance with work	State imposes a more severe penalty for non-compliance.	State reduces severity of worst sanction for non-compliance.
Duration of Worst Sanction	Length of most severe sanction for not complying with work requirements.	State increases duration of worst sanction.	State decreases duration of worst sanction.
Reapply	Whether the unit has to reapply after worst sanction for non-compliance has been imposed.	State adopts requirement that unit has to reapply after worst sanction has been imposed.	State abolishes requirement that unit has to reapply after worst sanction has been imposed.
Severity of Initial Sanction	Whether initial sanction for non-compliance is removal of full family benefit.	State adopts that full family benefit is removed as initial sanction.	State abolishes rule that full family benefit is removed as initial sanction.
<b>BROADER SET OF RULES:</b>			
<b><u>Earnings Disregards (1)</u></b>			
Earnings Disregards	Earned income an assistance unit can deduct without affecting benefits. Calculated for Month 5 for an individual working full-time at minimum wage.	Increase earnings disregarded for benefit calculation	Reduce earnings disregarded for benefit calculation
<b><u>Additional Eligibility Criteria (2)</u></b>			
Mandatory Job Search	Whether applicants are required to engage in job search activities before applying for benefits.	Adopt mandatory job search	Abolish mandatory job search
Diversion Payments	Whether the state offers a one-time payment if an eligible assistance unit does not apply for welfare for a certain period of time.	Adopt diversion payment	Abolish diversion payment

Source : Authors' calculations based on Urban Institute's Welfare Rules Database (Urban Institute, 2012).

Notes : See data appendix on further details how the policy rules were coded.

**Table A3: Summary Statistics**

	<u>Waiver Period</u>		<u>Tanf Period</u>	
	Mean	Std. Dev.	Mean	Std. Dev.
# Policy Experiments (Waiver Applications)	0.201	0.577		
Dummy for Policy Experiment (Waiver Application)	0.150	0.357		
# Policy Experiments (TANF period)			0.497	0.908
Dummy for Policy Reversal (TANF Period)			0.101	0.301
Governor's Past Vote Margin	8.497	7.316	11.42713	10.40595
Governor Lane Duck	0.238	0.426	0.303	0.460
Governor with Low Competence	0.491	0.500	0.462	0.499
Governor Age	52.840	7.856	55.616	7.576
Governor Party	0.587	0.489	0.063	0.318
Governor Party Ideology (Berry et al. 1998)	-0.019	0.279	0.063	0.318
Citizens Ideology (Berry et al. 1998)	47.400	15.210	51.282	15.989
Democratic Presidential Vote Share	42.725	9.050	47.620	9.925
Divided Government	0.555	0.497	0.547	0.498
Democratic Seat Share Upper House	0.598	0.185	0.512	0.162
Democratic Seat Share Lower House	0.589	0.179	0.519	0.154
Polarization Upper House	0.168	0.125	0.131	0.096
Polarization Lower House	0.162	0.118	0.125	0.091
Competitiveness Index (Holbrook and Van Dunk )	40.929	12.200	38.991	10.978
AFDC Spending (in US\$ millions)	330.4	683.9		
AFDC Spending (per Capita)	52.993	33.447		
TANF Spending (in US\$ millions)			490.430	990.560
TANF Spending (per Capita)			73.268	49.248
Unemployment Rate	6.466	2.110	5.280	1.848
Per Capita Income (/1000)	15719.2	5310.9	32342.4	7619.6
Population (/1000)	4.782	5.199	5.686	6.326
% Black Population	9.399	9.157	10.772	10.619
% Population 65+	11.941	2.272	12.722	1.805
% Immigrant Population	1.762	2.216	2.263	1.739
% Unmarried Births	24.096	8.812	35.518	7.428
Income Inequality (P90/P10)	7.895	1.741	8.833	1.992
Geographic Neighbors' Experiments (t-1)	0.220	0.409	0.341	0.452
Pop. Size Neighbors' Experiments (t-1)	0.203	0.351	0.501	0.658
Ideological Neighbors' Experiments (t-1)	0.229	0.393	0.492	0.660

*Notes:* For details on the welfare experimentation and reversal measures, see the data appendix. The quality of a governor is measured by years of political experience prior to becoming governor. The low quality governor indicator is equal to one if the number of years of political experience is below the median value. Governor Lane duck is equal to one if the governor cannot run for reelection. The past vote margin measures the winner's distance in votes to the runner-up in the last gubernatorial election. The ideology measure for the government is calculated from Berry et al. 1998 and ranges from zero for most conservative to 100 for most liberal, see the data appendix for details. The citizen ideology measure is also taken from Berry et al. 1998. Democratic Presidential Vote Share refers to the share in the last presidential election. Divided Government is equal to one if the party of the governor is different from the party of the majority of legislators in either the state's lower or upper house. The polarization variables are calculated as absolute deviations of the democratic seat share from 50%. The competitiveness index is the vote share of winning candidates in state legislative elections over the current and past 3 years. Geographic Neighbors' Experiments refers to the average number of TANF experiments for adjacent states (geographically neighboring states). Pop. Size Neighbors' Experiments refers to the average number of TANF experiments for states of similar population size (where all states are divided into ten bands of similar population size). Ideological Neighbors' Experiments refers to the average number of TANF experiments for the four states within the same census

**Table A4: Initial Reputation and Policy Experimentation (Linear Probability Model)**

Y: Applied for Waiver (Yes=1)	Reputation Effect				Governor Competence			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Past Governor Vote Margin	-0.0033* (0.002)	-0.0035* (0.002)	-0.0039* (0.002)	-0.0042* (0.002)		-0.0034* (0.002)		-0.0036** (0.002)
Lame Duck * Past Vote Margin			0.0020 (0.004)	0.0023 (0.004)				
Lame Duck			0.0082 (0.052)	0.0009 (0.054)			-0.0450 (0.030)	-0.0321 (0.032)
Governor with Low Competence					-0.0131 (0.031)	-0.0129 (0.031)	-0.0534 (0.035)	-0.0531 (0.035)
Lameduck * Governor with Low Competence							0.1582*** (0.048)	0.1543*** (0.048)
Year Fixed Effects	Yes	Yes						
State Fixed Effects	Yes	Yes						
Governor Age	Yes	Yes						
Change in Governor in Past Elections	No	Yes	No	Yes	No	No	No	No
Observations	785	785	785	785	872	872	872	872
R-squared	0.303	0.304	0.304	0.304	0.288	0.291	0.295	0.299

*Notes* : The dependent variable in all specifications is equal to one if a state has applied for a welfare waiver in a given year and zero otherwise. All specifications report estimates from a linear probability model. The specifications in columns (1)-(4) are the same as in Table 1; the specifications in columns (5)-(8) are the same as in Table 3. All specifications control for year and state fixed effects as well as governor age. The Columns (2) and (4) also include an indicator whether the governor has changed over the past two electoral cycles. Standard errors clustered at the state level are shown in parentheses. \*\*\* p<0.01, \*\* p<0.05 and \*p<0.1.

**Table A5: Alternative Measures of Experimentation in the TANF Period**

	<u>Restrict to First Change in</u> <u>Welfare Rule Only</u>			<u>Use Broader Set of</u> <u>Welfare Rules</u>		
	(1)	(2)	(3)	(4)	(5)	(6)
Past Governor Vote Margin	-0.0037** (0.001)	-0.0037** (0.001)	-0.0070*** (0.002)	-0.0025 (0.002)	-0.0025 (0.002)	-0.0048 (0.003)
Lame Duck * Past Vote Margin			0.0079 (0.005)			0.0045 (0.006)
Lame Duck			-0.0905 (0.088)			-0.0150 (0.103)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Change in Governor in Past Elections	No	Yes	Yes	No	Yes	Yes
Observations	544	544	544	544	544	544
R Squared	0.226	0.226	0.231	0.233	0.234	0.236

*Notes* : The table reports the baseline from Table 2 using alternative measures of policy experimentation during the TANF period. The first measure in columns (1)-(3) uses the same set of rules, but focuses on the first change in each welfare rule. The second alternative measure in columns (4)-(6) uses a broader set of welfare rules (including earnings disregards, mandatory job search and diversion payments) to define an experiment. See the data appendix and Table A2 for details of the coding. See notes to Table 1 for details on the specification. Standard errors clustered at the state level are reported in parentheses. \*\*\* p<0.01, \*\* p<0.05 and \* p<0.1.

**Table A6: Governor Quality and Policy Experimentation during the TANF Period**

	<u>Main Effect</u>		<u>Differential Effect</u>		<u>Control for</u>	
	<u>Competence</u>		<u>for Lame Ducks</u>		<u>Potential Gains</u>	
	(1)	(2)	(3)	(4)	(5)	(6)
Governor with Low Competence	-0.0574 (0.071)	-0.0564 (0.071)	-0.0993 (0.084)	-0.0725 (0.090)	0.0829 (0.102)	0.0806 (0.102)
Lame Duck * Governor with Low Competence			0.0595 (0.128)	0.0378 (0.134)	-0.0841 (0.128)	-0.0804 (0.127)
Lame Duck			-0.0264 (0.078)	0.0031 (0.082)	0.0180 (0.077)	0.0481 (0.081)
Past Governor Vote Margin		-0.0035 (0.003)		-0.0038 (0.003)		-0.0053** (0.002)
Past TANF Spending					0.0003* (0.000)	0.0003* (0.000)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Governor Age	Yes	Yes	Yes	Yes	Yes	Yes
Observations	682	682	700	682	539	539
R-squared	0.413	0.412	0.411	0.414	0.338	0.343

*Notes* : The dependent variable in all specifications is the number of policy experiments in the TANF period in a given state and year. Columns (1) and (2) includes an indicator whether a governor is below the median political experience in the governor sample (as a proxy for the governor's competence). Columns (3) and (4) allow the effect of governor competence to differ for governors who cannot be reelected ("lame ducks"). Columns (5) and (6) also control for past TANF spending (measured in US\$ millions) as a proxy for potential gains from experimentation. All specifications control for year and state fixed effects as well as governor age. Even columns add the governor's vote margin in the last gubernatorial election (measured as the distance to the runner-up in percent). Standard errors clustered at the state level are shown in parentheses. \*\*\* p<0.01, \*\* p<0.05 and \*p<0.1.