

Trial and Error?

Policy Experimentation during the US Welfare Reform

Andreas Bernecker* and Christina Gathmann†

Preliminary – Comments welcome

This version: March 2013

First version: April 2012

Abstract

Sweeping reforms often start out as experiments and many reforms are first tested locally before they are implemented nationwide. Yet, little is known empirically whether political incentives facilitate or impede policy innovation; and whether policy-makers reverse policies that turn out to be unsuccessful. In this paper, we study whether reelection concerns affect policy experimentation and learning in the US states. Our empirical setting is the 1996 Welfare Reform, which has shifted the autonomy to implement welfare policies from the federal government to the individual states. Based on a model of reputational concerns, our empirical results suggest that reelection concerns matter both for policy experimentation and policy reversal.

Key words: Policy Innovation, Experimentation, Reputation, Welfare Reform

JEL codes: D72, D78, H11, H53, I38

Acknowledgments: We thank Toke Aidt, Manuel Arellano, Pierre Boyer, Matz Dahlberg, Peter Fredriksson, Jörg Oechssler, Per Pettersson-Lidbom, Federico Revelli, Elisabeth Schulte-Runne, Christoph Vanberg and seminar participants at Heidelberg, the Silvaplane Workshop on Political Economy, and the Evaluation of Political Reforms Conference in Mannheim for helpful suggestions. We are very grateful to Greg Shaw and Dave Andersen for sharing 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 detailed questions. Jan Scherer, Michael Hellwig, and Robert Aue provided outstanding research assistance. We are responsible for all remaining errors.

*Andreas Bernecker, Department of Economics and Collaborative Research Center "Political Economy of Reforms" (SFB 884), University of Mannheim, Germany, andreas.bernecker@gess.uni-mannheim.de.

†Corresponding author: Christina Gathmann, Department of Economics, University of Heidelberg, IZA and CESifo, christina.gathmann@awi.uni-heidelberg.de.

1 Introduction

When President Clinton signed the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) in August of 1996, it marked the most dramatic reform of US welfare policy since the New Deal. The law abolished the entitlement to cash assistance for poor families, the “Aid to Families with Dependent Children” (AFDC). The newly established “Temporary Aid to Needy Families” (TANF) defined welfare benefits as temporary assistance rather than an entitlement. Also, TANF effectively decentralized the authority to design welfare policy to state policy-makers. After August 1996, states were free to choose their own eligibility criteria for benefit receipt, which requirements recipients had to fulfill and what sanctions to impose if recipients did not comply with the requirements.¹

A vast literature has analyzed the consequences of the 1996 welfare reform for caseloads, employment, earnings, poverty and many other outcomes (see, for example, Blank (2002); Grogger and Karoly (2005); and Moffitt (2001) for excellent surveys). Little is known, however, how state governments actually chose their welfare policies (but see Soss et al. (2001)).² Many of the new welfare rules that were eventually adopted had never been implemented before. As a consequence, states had little prior experience what policy rules worked and which combination of rules worked best in their particular circumstances.

The question we ask in this paper is what determined whether and how much states experimented with new welfare rules? We argue that reelection concerns played an important role for the decision to experiment and possibly reverse welfare rules later on. To study the role of electoral concerns for experimentation, we build a theoretical model of reputation concerns in which both voters and politicians are uncertain about the best welfare policy (along the lines of Majumdar and Mukand (2004)). The basic intuition of the model is the following: a politician can either stick to an existing policy or implement a new policy that may be better or worse than the status quo (but is strictly better when the new policy is appropriate for the current situation). Politicians differ

¹The federal government only provides some rules how federal funds can be spent. For example, 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) have to use state funds to support recipients for more than 60 months. States could however also decide to implement shorter lifetime time limits on benefit receipts. We return to a more detailed discussion of the welfare policy rules below.

²In contrast, there are several good studies of the political process leading up to the TANF reform at the federal level (Reintsma (2007); Weaver (2000); or Haskins (2006)).

in their quality, i.e. some politicians are better able than others to identify policies that fit the state's circumstances. Voters however, do not know the quality of a politician for sure and therefore infer her quality from the observed outcomes of the policy. Suppose that a high-quality politician always implements the right policy: she will experiment with the new policy if it is appropriate; she will not experiment if the status quo policy is better. A high-quality politician will never reverse a policy experiment even if the performance of the policy is disappointing at first. A low-quality politician in contrast, will try to mimick the high-quality politician in order to increase her reelection chances. As a consequence, a low-quality politician might experiment to make voters' believe she is a high-quality politician. A low-quality politician will also be hesitant to reverse a policy experiment because of its high reputational costs (as it signals to the voter that the politician is indeed low-quality). We derive a number of empirical predictions how a politician's quality and initial reputation affect her willingness to experiment with a new welfare policy and to possibly reverse it.

To test our predictions in the context of the US welfare reform, we collect a new dataset on the welfare policies states adopted under AFDC and TANF. We complement the panel of policy rules with additional information on the political process, the characteristics of the governor and measures of the socioeconomic conditions in each state.

Our results suggest that reputational concerns matter both for policy experimentation and policy reversals during the welfare reform. In particular, we find that high-quality governors are more likely to experiment the higher the potential net gain from experimentation (increasing her propensity of experimentation by 37%). Low-quality governors in contrast, are less likely to reverse a policy when the net gains are high (a decrease in the propensity to reverse by 36%). We also show that higher initial reputation of the politician reduces policy experimentation by about 27%. A number of robustness checks bolster our findings that reelection concerns matter. We find little evidence, for example, that the ideology or partisan composition of state governments alone explains the pattern of policy experimentation and policy reversals we observe in the data. The policy experiments observed during the US welfare reform support the idea that electoral incentives (and reputation-building by the governor) are important for understanding policy innovations.

Our empirical analysis contributes to a large literature in political science studying

policy innovation and diffusion in multiple contexts (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 good surveys). Our analysis differs from these prior studies along at least two dimensions: first, we analyze the political economy of policy innovation focusing on the role of reelection concerns. Second, we analyze for the first time policy experiments implemented during the US welfare reform, the most dramatic shift in welfare policy since the New Deal.

Hence, we provide systematic quantitative evidence where there have been rich case studies about the process of welfare reform in individual states (Francis (1998); Tweedie (2000) or Weissert (2000), for example) and the welfare devolution (Soss et al. (2001); Lieberman and Shaw (2000)). Further, our analysis also contributes to a small empirical literature studying the role of electoral concerns for tax and expenditures (see Besley and Case (1995b), Besley and Case (1995a) or List and Sturm (2006), for examples).

Our theoretical setup is related to a large literature on career or reputational concerns following the seminal work of Holmström (Holmström (1982); Holmström (1999)). An early application to the political arena is the political agency model (Rogoff (1990)). Closer to our setting are models where voters and politicians are both uncertain about the mapping from chosen policies and realized outcomes (Harrington (1993); Majumdar and Mukand (2004); or Fu and Li (2010)). As in our setting, politicians have an incentive to manipulate voters' beliefs about the politician's ability to enhance their reelection chances. Our study provides the first empirical test of these predictions for policy innovation.

An alternative explanation for policy experimentation relies on the politician's ideology. One important contribution (Cukierman and Tommasi (1998)) shows, for example, that politicians can boost their reputation among voters the more atypical their policy choice given their ideology.³ We show in the empirical analysis that ideology alone is unlikely to explain the policy experiments we observe during the US welfare reform.

Our study is also closely related to the literature on the question why certain policies persist (even if they are known to be a failure). One reason why welfare-improving reforms might not be implemented is individual-specific uncertainty about who wins and

³Moen and Riis (2010) relax some restrictive assumptions of Cukierman and Tommasi (1998) and show that atypical policies might still emerge in equilibrium.

who loses from the proposed changes (Fernandez and Rodrik (1991)).⁴ Reforms might also be strategically delayed by politicians in a war of attrition to shift the burden of a reform to supporters of political opponents (Alesina and Drazen (1991)). And inefficient policies could also persist if vested interests emerge when individuals strategically adapt to status quo policies (Coate and Morris (1999)). We focus instead on the reputational costs of policy reversals as one explanation why policies might persist.

The next section introduces background information about the US welfare reform and discusses the role of state governors during the reform process. In Section 3, we discuss the theoretical model and how reputational concerns encourage or inhibit experimentation and the reversal of potentially failed policies. We describe our data sources we collected in Section 4. Section 5 presents the results of our basic predictions and several robustness checks. Finally, Section 6 concludes.

2 The 1996 Welfare Reform in the United States

2.1 From AFDC to TANF

The passage of the Personal Responsibility and Work Opportunity Reconciliation Act (PROWRA) in 1996 abolished the old federal welfare program, the “Aid to Families with Dependent Children” (AFDC), and replaced it with the “Temporary Assistance to Needy Families” (TANF) program.

The existing AFDC was an entitlement program that provided financial assistance for any family meeting the eligibility requirements. Federal rules determined the eligibility criteria and provisions under AFDC. In principle, all single-parent families with dependent children were eligible for aid.⁵ There was little room left for states to design welfare policy under AFDC. States could implement additional eligibility rules: several states passed ‘fit parent’ or ‘suitable home’ provisions to limit payments to families with unsatisfactory behavior, for example. In addition, states could choose the welfare benefit level under AFDC which varied (for a family of three with no other income) from \$119

⁴Jain and Mukand (2003) extend the framework by allowing for compensatory transfers and show that still there might be a non-monotonic relationship between the expected welfare gains of a reform and its likelihood of political implementation.

⁵Although a separate program for jobless two-parent families (AFDC-UP) was established in 1961, single-parent families remained the primary beneficiaries of the AFDC program prior to 1996.

in Mississippi to \$720 in Alaska in 1995.

Since 1962, section 1115 of the Social Security Act granted authority to the Secretary of Health and Human Services to waive the federal rules and regulations governing AFDC. Under this provision, states could petition the U.S. Department of Health and Human Services (DHHS) to implement experimental, pilot, or demonstration projects.⁶ To apply for a waiver, states submitted detailed requests which rules and program elements were modified and what new rules and regulations were implemented. The proposed waiver provisions were then reviewed, and sometimes modified, 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 either as is or after some changes requested by DHHS were made to the application. In our empirical analysis, we will use the number of waivers a state applied for - irrespective of whether the waiver was approved, amended, withdrawn or denied - as one measure of policy experimentation.

Welfare waivers were rare until the late 1980s when AFDC caseloads approached almost 4 million families or around 11 million recipients nationwide (see Figure 1 for the evolution of caseloads in the whole country and some selected states). Waivers became a popular policy instrument to reduce state caseloads when federal aid for AFDC was cut during the Reagan years and funds became tight 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 representing 43 states and the District of Columbia. In total, all but five states received approval for one or more waivers.

Most waivers implemented multiple changes to welfare rules simultaneously. Most popular were experiments to allow for more generous earnings disregard (thus reducing work disincentives), but also to impose stricter work requirements (to increase employment), time limits on benefits, sanctions (for families that failed to meet work requirements), and family caps (eliminating or reducing the benefit increase if a child is conceived

⁶These experiments were required to be cost-neutral and to include a rigorous evaluation of the project.

during welfare receipt). The data appendix provides more detailed information on these novel policies which were first implemented during the waiver period.

With the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) in August of 1996, the new TANF program effectively decentralized the authority to design welfare policy (within the limits of federal guidelines) to the individual states.⁷ After August 1996, states were free to choose their own eligibility criteria, work requirements, earnings disregards, what sanctions to impose if recipients failed to meet the work requirements and many other rules. States could, for example, impose time limits that were stricter than the federal limit of 60 months of benefit receipt over a person's lifetime; or, they could decide instead to use state resources to extend the time limit beyond 60 months (or even impose no lifetime limit for benefit receipt at all). Because states had to decide on many new policy dimensions, welfare rules in the states have changed gradually even after 1996 and continue to do so until today.

2.2 The Political Economy of Welfare Reform

Rising AFDC caseloads and welfare dependence were hotly debated political issues in the mid and late 1980s. A prominent role in pushing welfare reform and welfare waivers on the political agenda was hereby played by state governors. Wisconsin's Republican governor Tommy Thompson, a leading state figure of welfare, is 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. Under Thompson's leadership and using his line item veto power, the new Wisconsin Works, better known as W-2, was implemented in 1993. With its focus on labor market participation with stricter work requirements and harsh sanctions for noncompliance, the W-2 program became a role model 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))

⁷Waiver provisions remained in place even after the passage of PRWORA until the waiver expired or was replaced by new state welfare rules.

or Winston (2002) for descriptive case studies).⁸ Given the governor's prominent role, we next outline a model stressing the decision-making governor as the key player during the welfare reform.

3 A Model of Reputational Concerns [Preliminary]

To study the decision to experiment with (and possibly reverse) welfare policies in the US states, we discuss a simple model of reputation concerns and highlight its empirical predictions. We focus on a reputation model of a single political decision-maker who is concerned both about welfare and reelection (following the setup of Majumdar and Mukand (2004)).

3.1 Basic Setup

The incumbent politician can choose between keeping the current policy or implementing a new policy. The return from the status quo policy is public information but both voters and politicians are uncertain about the returns to the new policy. The returns from the innovative policy may be higher or lower than the status quo: Social welfare will be higher if the innovative policy is the appropriate policy for the state's particular circumstances.

Politicians differ in their quality which is private information. High-quality politicians are better able than others to identify and implement policies that best fit the current state of the world. Voters do not know whether the observed outcome of the new policy is the result of a high-quality politician (making the right policy choice) or of a low-quality politician whose policy choice turns out to be successful. After the first period, the incumbent can decide to revert back to the old status quo policy (at no cost) or stay with the implemented policy experiment. Because politicians care about their reelection as well as about welfare, they might experiment with new policies for strategic reasons, i.e. to make voters believe that they are of high quality.

Suppose first the politician only cares about social welfare (i.e. the outcome of the chosen policy). In this first best case, the decision to experiment with a new policy and the

⁸Another indication for the role of state governor is that 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)).

decision to reverse a previously implemented policy depend on the expected net welfare gain alone. Hence, the decision to experiment will depend on the costs of implementing or reversing the policy (e.g. administrative costs) and the expected welfare gain from experimentation or reversal. For policy reversals, the decision will in addition depend on the observed outcome of the policy experiment.

Majumdar and Mukand (Majumdar and Mukand (2004)) demonstrate that there is a unique Bayes Nash equilibrium characterized by: (a) two cutoff rules (whether a politician experiments or not; and whether she reverts back to the status quo policy or persists with the experiment); (b) a set of voters' beliefs about the politician's quality after observing the policy choice and outcome; and (c) the reelection probability for the incumbent politician given voters' beliefs.

In equilibrium, politicians sort themselves into three groups: those that do not experiment, those that do experiment but later revert to the status quo and those that do experiment and stick to the policy experiment. Uniqueness of the sorting equilibrium is achieved by three specific features of the model: first, there are only two types of politicians (high-quality and low-quality) and the high-quality politician is perfectly informed about the appropriate policy. Second, the analysis studies electoral incentives only within a single electoral cycle. And third, the cutoff rule that defines the decision to revert or persist with a policy experiment is independent of the cutoff rule that defines whether to experiment or not.⁹

3.2 The Role of Reelection Concerns

Suppose now that politicians also care about their reelection, for example, because they derive some ego rents from being in office. Voters want to reelect high quality politicians who are better able to identify and implement the policy that is most appropriate for the current state of the world. Voters observe the chosen policy and a noisy signal of the outcome of policy decisions; they use this information to update their beliefs about the politician's ability. The politician then has an incentive to strategically manipulate voters' beliefs in order to increase her reelection chances.

Take as example a high-quality politician who gets a perfect signal about the state of

⁹The model can be generalized to multiple types as Majumard and Mukand show in an earlier working paper version. This more general model yields similar predictions than the baseline setup discussed here.

the world (and hence, knows what is the appropriate policy to choose). A high-quality politician will therefore innovate if the new policy is appropriate (the policy that produces the largest expected net benefit). Also, she will always persist with a policy experiment (even if the observed outcome fails to be realized) because she knows that the policy experiment is the right one for the current state of the world.

A low-quality politician instead might try to manipulate voters' beliefs by mimicking the high-quality politician. She might choose to experiment even when the expected net benefit of the new policy are low because the reputational gains of experimenting are large if it is successful (i.e. the appropriate answer to the current, but uncertain state of the world). In that case, voters will update their beliefs that the politician is of high quality. However, we also might get too little experimentation from low-quality politicians. If the expected net gains from experimentation are high, experimentation (among all politicians) becomes more likely. In that case, the reputational gains from experimentation are low, reducing the low quality politician's incentive to implement a policy experiment.

Strategic considerations also influence the decision to revert back to the status quo (when the outcome of a policy experiment has been publicly observed). Reputation concerns might induce a low-quality politician to inefficiently persist with an experiment even if the policy is not appropriate for the current state of the world. The reason is that a policy reversal would signal to the voter that the politician is of low quality (because a high-quality politician never reverses a policy). If reputation concerns are sufficiently important (relative to welfare considerations), the politician experiments either too much or too little; she might also be hesitant to reverse a policy.

The propensity to experiment and reverse a policy experiment also depends on the politician's initial reputation. The higher her initial reputation, the fewer are the potential gains from experimenting in order to manipulate voter beliefs. Hence, a high initial reputation should make reckless experimentation less likely. Similarly, a high initial reputation makes it less likely that a policy is reversed because the reputational costs of policy reversals are high.

The model also can incorporate governor ideology and analyze its role for policy experimentation and reversal decisions. To preserve analytical tractability, ideology enters the model in a very specific way: ideology influences the politician's initial belief whether

the new policy is appropriate for the current state of the world. A status quo oriented politician believes that the current policy is the most appropriate for the state of the world. At first sight, a status quo oriented politician should be less likely to experiment (because she believes the innovative policy is not the appropriate policy). However, reputation concerns might reverse this initial intuition. A status quo oriented politician has higher reputational gains if she experiments with a policy that turns out to be successful (compared to a politician who believes the new policy is the appropriate one). If reputation concerns are strong enough, the status quo oriented politician is actually more likely to experiment - contrary to her own ideology. A similar argument applies to policy reversals: a status quo oriented politician faces higher reputation costs if she reverses an innovative policy (compared to a politician who believes the new policy is appropriate for the current state of the world); she might therefore be less likely to reverse a policy experiment.

3.3 Empirical Predictions

We can derive a number of empirical predictions from the reputational model sketched above. Consider first the effects of politician quality (as the key variable of the reputational model) and the expected net welfare gain of the policy choice. The first prediction studies policy experimentation. As shown in the equation below, the propensity to experiment depends on the politician's quality, on the expected net welfare gain of the experiment, and on the interaction of both. The key parameter of interest is δ which we expect to be positive. A high quality incumbent can be expected to experiment more when the expected net gains from doing so are large. A low quality politician is less likely to experiment with large benefits because the reputational gains are comparatively low.

$$Pr(Experiment) = \alpha + \beta * Quality + \gamma * Netgain + \delta * Quality * Netgain \quad (1)$$

We next study the determinants of policy reversals in the model. Here, we are particularly interested in low quality politicians where strategic considerations are likely to be important. The probability to reverse a new policy previously implemented depends in the model on the quality of the politician, on the expected net gain, and the interaction

of the two. The key parameter in the equation below is again δ which we expect to be negative. Low quality politicians are unwilling to revert a policy especially when the expected net gain from an experimental policy is large. The intuition is that the reputational gain of sticking with the innovative policy are especially large when the policy is successful in the second period.

$$Pr(Reversal) = \alpha + \beta * Low_Quality + \gamma * Netgain + \delta * Low_Quality * Netgain \quad (2)$$

We next discuss the role of the politician’s initial reputation as perceived by the voter. In the model, politicians with high perceived quality (reputation) have higher probability of being reelected compared to a politician with lower reputation. Therefore, a politician with high reputation has no need to engage in risky policy experiments (to boost his reputation). A politician with low reputation in contrast, might experiment recklessly in the hope of grabbing large reputational gains. Hence, we expect the probability of policy experimentation to decrease with the incumbent’s reputation (a negative β in the equation below).

$$Pr(Experiment) = \alpha + \beta * Initial_Reputation \quad (3)$$

In the robustness section, we also return to the question how a politician’s ideology affect policy experimentation and reversals.

4 Data Sources

We collect a comprehensive new dataset of state-level welfare policy rules, political conditions and socio-economic characteristics. Here, we summarize our key sources and variables to study policy experiments and reversals during the US welfare reform. The data appendix provides a more detailed description of all data sources and the construction of key variables.

Policy Experimentation and Reversals

To measure the extent of policy experimentation, we use the waivers states applied for at the Secretary of Health and Human Services (HHS). We first code the date a state applied

for a waiver from Koerper (1996), Crouse (1999) and Lieberman and Shaw (2000). We include waiver applications that have been approved and implemented, but also those that were denied or withdrawn by the state. We use information on all waiver applications irrespective of whether the state planned to implement it only in specific counties or the whole state. Altogether we have information on 195 waivers states applied for between 1978 and 1996. While US state governments could apply for waivers since 1962, states only began to use waivers starting in the late 1970s. Waiver activity stopped with the passage of the TANF reform at the federal level in 1996. Figure 2 shows the evolution of waiver activity from 1978 to 1996. Table 1 goes into greater detail and shows the distribution also across states. The last column shows the average number of waivers states applied for in each year.

Our first measure of policy experimentation 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. Most states only apply for one waiver within a year. In twenty cases, a state applied for two waivers within the same year, in seven cases for three waivers and in two cases for four waivers (see Table 1). The maximum number was filed by California and Illinois which applied for six waivers each in 1994 and 1992 respectively.

To measure policy reversals and policy experimentation after 1996, we collected the welfare policy rules in each state and year since 1996 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 over the period from 1996 to 2010. The database contains literally hundreds of rules on eligibility, benefit calculation and many other aspects of welfare reform. We focus on a subset of policy rules that were at the center of the public and academic debate surrounding the US welfare reform. Specifically, we code rules in four main policy areas: (a) whether a state has implemented a family cap; (b) the time limit a state adopted; (c) the work requirements a state requires; and (d) when and how severe sanctions are applied in case an adult recipient does not adhere to the work requirements. Table A1 provides a detailed explanation about the individual policy rules and how we define a policy experiment or reversal in each case.¹⁰

¹⁰Reputational gains from experimentation should be largest for policy rules that are salient during the reform period. To test the sensitivity of our results to the selection of policy rules, we construct reversal

Based on the policy rules, we define a policy reversal if a state reversed any rule (in one of the policy areas above) in the post-TANF period that had been adopted by the state by 1996 or thereafter. The variable is a binary indicator which is equal to zero if no policy reversal took place in a given year. Take, for example, family caps which require that benefit payments do not increase if the recipient family conceives an additional child while receiving welfare. While family caps did not exist under AFDC, twenty-four states adopted a family cap between 1996 and 2010. Several other states had already adopted a family cap during the waiver period. Over the 1996 to 2010 period, eight states also abolished the family cap they had adopted before. We then 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 (see Table A1 for further details).¹¹ As an alternative outcome measure, we use the number of policy reversals in a given year. Figure 3 shows the evolution of policy reversals over the period from 1996 to 2010. Overall, we observe 109 policy reversals (in 14% of the state-year observations). In most cases (N=85), states reverse a single policy rule in a given year (in 20 cases, states reversed two rules and in 4 cases, three policy rules in the same year).

Electoral Incentives

Our model of reputational concerns emphasizes two characteristics of the governor: the quality of the governor and her initial reputation. Both quality and initial reputation are difficult to measure as they are ultimately unobservable. To proxy for governor quality, we use the years a governor has spent in all prior political offices (see the data appendix for details). Cumulative years in political offices will be a good indicator of government quality if elections (or a sequence of elections) select the most able candidate and/or candidates learn about policy-making from spending time in political office. While the number of years a governor has spent in the current office is in principle observable

measures based on a broader set of rules (covering demographic and financial eligibility requirements, schooling, health and immunization requirements for continued eligibility and eligibility rules of non-citizens, for example). Our results are robust with respect to these sensitivity tests.

¹¹Note that we can measure the adoption and abolishing of policy rules consistently only for the years since 1996 when the policy rules appear in the Welfare Rules Database of the Urban Institute. Changes in policy rules during the waiver period are not captured by our measure 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. Anecdotal evidence suggests that these policy reversals were rare in the pre-TANF period (many waivers only expired after the federal TANF reform in 1996 and many states adopted their own TANF policies only after 1996). Our policy reversal measure thus, if anything, understates the number of actual policies reversed by state governments.

to voters, most voters do not know about the total political experience a governor has accumulated over her career.¹²

The second important characteristic is the governor’s initial reputation. We use the vote margin in the past gubernatorial race as an indicator of reputation (from List and Sturm (2006) and Leip (2012)). Governors which are thought to be capable to choose the right policies are likely to be elected with a larger majority than governors that just barely reach a majority of votes.¹³

Reputational concerns should also be stronger among governors that can be reelected compared to governors that face binding term limits (“lameduck”). Information on gubernatorial elections and term limits are taken from multiple sources (List and Sturm (2006); Council of State Governments (2012); and Leip (2012)).

To analyze the role of ideology, we use two measures: the party of the governor and a measure of governor ideology (based on Berry et al. (1998)). The latter variable takes into account differences in ideology between different governors of the same party; the measure varies between 0 and 100 where larger values represent a more liberal attitude of the governor.

Whether policy experiments (and reversals) are undertaken also depends on the net gain of an experiment. In our setting, states experiment to learn about the effects of alternative welfare policy rules. Given the high caseloads that were reached prior to the reform, an important component of the gain from experimenting is how much it could reduce welfare caseloads (after the TANF reform, state governments only received block grants from the government and had to finance excess caseloads out of the state budget). We therefore use the lagged caseload per capita in a state as an indicator of the potential gain of an experiment.

¹²Formal educational qualifications might be an alternative measure (used in Galasso and Nannicini (2011) and Gagliarducci and Nannicini (2013) for Italian politicians or Kotakorpi and Poutvaara (2011) for Finnish MPs. In our analysis, educational qualifications were never found to be a statistically or economically relevant determinant of policy experimentation. One explanation is that we only have crude measures of formal education (no college degree, BA or equivalent or Master degree and other postgraduate degrees) which might not contain enough variation to identify an effect. The substantive explanation is related to the fact that most of our governors 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 has a highschool degree. Even in the Finnish parliament, only about 20-30% of the MPs have a master degree or more. It could therefore be that formal educational credentials are a prerequisite to become a state governor, but do not act as a signal of quality.

¹³One problem with this measure of reputation is that a small vote margin also depends on the quality of the challenger in the past election. As an alternative measure, we use governor’s job approval ratings when elected into office as a proxy for her initial reputation.

Other Political Variables and Demographics

To control for other political forces and socio-economic conditions in a state, we collected a number of additional variables. To control for the influence of the legislature (both lower and upper chamber), we collected data on the party composition and political polarization (Klarner (2003)). Polarization is measured as the Democratic seat share in the lower (upper) chamber calculated as the deviation from 50%. We further coded whether the government in the state is politically divided between Democrats and Republicans, i.e. whether the party of the ruling governor is different from the majority in the state parliament (also from Klarner (2003)). To control for socioeconomic conditions and voter preferences in a state, we also collect information on personal income per capita, the size of the population, the share of the Black population in a state and the share of the population aged 65 or above (all from United States Census Bureau (2011)) as well as income inequality (authors' calculation from the Current Population Survey (Center for Economic and Policy Research (2012))) and the size of the immigrant population (U. S. Department of Homeland Security (2011)). To control for citizen ideology, we use the percentage of votes for a Democratic candidate in the last presidential election (United States Census Bureau (2011)). Table 2 contains summary statistics of the variables used in our analysis separately for the period of waiver experimentation (1978-1996) and the post-TANF reform period (1996-2010).

5 Empirical Results

5.1 The Role of Quality and Reputation

The first prediction from the model concerns the role of quality. High-quality governors who receive valuable information about the state of the world are likely to experiment more when the net gains from experimentation are high. Table 3 tests this prediction regressing the number of waiver applications in the pre-TANF period on the quality of a governor, the potential net gain (caseload) and the interaction between the two.¹⁴ Specification (1) controls for year fixed effects, specification (2) adds state fixed effects,

¹⁴Since the number of waivers is a count variable truncated at zero and with integer increments, a Poisson model might be more appropriate. Our first tests show that results are similar when using a Poisson model instead.

specification (3) adds the age of the governor as a potentially relevant control. In line with the model, we find that the propensity to apply for welfare waivers increases in the quality of the governor when the gains from experimenting are high. The coefficient is economically and statistically significant in all specifications. An increase in the net gain and quality of the governor by one standard deviation increases the propensity to experiment by 37.30 %.¹⁵

Note that reputational concerns about reelection prospects (assuming that voters are more likely to reelect a politician they perceive as more competent) should be stronger for governors that can stand for reelection. Reputational concerns should be weaker (or even fully absent if the politician is not interested in another political office) for lameducks, i.e. governors that face binding term limits. We test whether quality matters more for policy experimentation among governors that can be reelected in specifications (4) to (6). The results show that the propensity to experiment increases much more in quality and the net gains from experimentation when a governor can be reelected. The effect is even negative and typically not statistically significant for lameducks, i.e. governors that have limited (or no) electoral incentives to experiment. In contrast, the effect for governors who can be reelected even gains in size and significance in specifications (4) to (6). Table 4 tests the same prediction for the period after the TANF welfare reform (1996 to 2010). Again, there is a statistically significant effect in the expected direction for governors who can be reelected, but not for lame ducks.¹⁶ This provides first evidence that reputational concerns matter for policy experimentation.

An additional prediction of the model relates policy experimentation to a governor's initial reputation. A high initial reputation reduces the propensity to experiment as the reputational costs of implementing a policy experiment that fails are higher. Table 5 tests this prediction and generally finds evidence that is consistent with the model's prediction. Again, specification (1) includes year fixed effects, specification (2) adds state fixed effects, and specification (3) adds the age of the governor. Specifications (4) to (6)

¹⁵Take specification (6). One standard deviation (0.43) of quality*caseload multiplied with the estimated coefficient (0.586) yields an effect of 0.25. The effect of quality itself is also significant. One standard deviation of quality (9.09) multiplied with the estimated effect (-0.0195) yields an effect of -0.18. The sum of both effects (0.25-0.18) evaluated at the mean number of experiments (0.2) yields an effect of 37.30%.

¹⁶The effect is 49.43% and thus of roughly similar size as before. Take specification (6). One standard deviation (0.284) of quality*caseload multiplied with the estimated coefficient (1.671) yields an effect of 0.475. Evaluation at the mean number of experiments (0.96) yields an effect of 49.43%.

differentiate the effect for lame ducks versus governors who can be reelected. Indeed, the propensity to experiment declines with a governor's initial reputation. The effect is again sizeable: Increasing initial reputation (the vote margin in the last election) by one standard deviation reduces the propensity to experiment by 26.93%¹⁷. The effect is again more pronounced for governors who can be reelected. It even flips sign for lame ducks.

5.2 Quality and Policy Reversals

In the model, governor quality also has an effect on the propensity to reverse a potentially failed policy. Low-quality governors face high reputational costs of reversing a policy experiment. They might be especially unlikely to reverse a failed policy when the expected net gain from the experimental policy is high. This is because the gains for the politician from being lucky with the policy in the second try are high. Table 6 regresses the indicator of observed policy reversals on the net gain (measured by caseload as before), whether a governor is of low quality (has number of years in political office below the median) and the interaction between the two variables. As in the analysis of policy experimentation in Tables 3 and 4, specifications (1) to (3) show a baseline, specifications (4) to (6) control for lame ducks. Based on our reputational model, we expect the coefficient on the interaction effect to be negative. The results confirm that prediction: low quality governors are indeed less likely to reverse a policy experiment when the net gains are high. The size of the effect is around 35.63%, i.e. a one standard deviation increase in net gain among low quality governors reduces the chance of a policy reversal by about 35.63%.¹⁸

This result suggests that reputation not only plays a prominent role for whether a state experimented with new welfare policies, but also for later reversal decisions during the TANF period.

¹⁷Take specification (6). One standard deviation (7.25) of the past vote margin multiplied with the estimated coefficient (-0.00743) yields an effect of -0.0539. Evaluation at the mean number of experiments (0.2) yields an effect of -26.93%.

¹⁸Take specification (6). One standard deviation (0.0115) of $lowquality * caseload$ multiplied with the estimated coefficient (-4.74) yields an effect of -0.055. Evaluation at the mean number of reversals (0.153) yields an effect of -35.63%.

5.3 Additional Tests and Robustness Checks

We check whether other political forces could explain our results regarding policy experimentation. First, we take into account the potential role of political ideology for policy experimentation decisions. Table 7 basically shows the baseline from Table 3 (analyzing the effect of quality on experimentation), but adds several ideology measures as further controls. Specification (1) is to illustrate that it is not a very simple governor ideology story driving policy experimentation: Republican governors do not react significantly different to rising welfare caseloads than Democratic governors. Specifications (2) to (6) present the reputational channel from before, but control for the governor party, the ideology of the governor party, the ideology of citizens, and the Democratic vote share in the last presidential election. We find that none of these ideology measures matters and that our results from before are unaffected.

Further, our model of reputational concerns features a single decision-maker, the state governor. State governors played a key role during the waiver period and the implementation of TANF (see section 2 above). And yet, the legislature has certainly a say in welfare policy-making as well. Table 8 tests whether reputational concerns of the governor still matter for policy experimentation and reversals if we include political controls. We find that neither the composition of the legislature (upper and lower house) nor the polarization in the legislature (often seen as an indicator for stiff political competition) has any effect on the extent of policy experimentation or policy reversals. We test whether a divided government (the party of the governor differs from the majority in the legislature) has any effect as divided governments are often thought to take longer to reach decisions (Besley and Case (2003)). We find that states with divided governments actually experiment more.¹⁹ But again, our conclusion that reputation concerns of the governor matter for policy experimentation is not impaired. Overall, these political robustness checks support our basic story that reputational concerns of the governor are an important motive for policy experimentation.

An alternative explanation could be that policy experimentation is influenced by actions of neighboring states (as suggested by models of yardstick competition (Besley and Case (1995b)) or other strategic or informational spillovers across states (as stressed by

¹⁹For a detailed analysis of the effects of divided government on reform-making in US states, see Bernecker (2013).

the many studies of policy innovation in the political science literature). Table 9 analyzes this issue. As a first empirical test, we check whether the experimentation in neighboring states (measured as the mean amount of experimentation among a state’s geographic neighbors) influences the propensity to experiment. Specifications (1) to (3) show that neighbors’ experimentation does not significantly affect a state’s propensity to experiment. Also, the estimated effect of the reputational channel is unaffected. Specifications (4) to (6) conduct a similar test, but this time focusing on states with a similar population size (instead of geographic neighbors). Again, we do not find evidence for spillover effects.

Finally, Table 10 includes several proxies for voter preferences in a state. Specification (1) adds the unemployment rate, the per capita income, and the population size as controls. Specification (2) further adds the share of the Black population and the share of the population that is 65 or older. Specification (3), finally, additionally includes the share of immigrants and the Gini coefficient of family income as a measure of inequality.²⁰ All specifications again also include state and year fixed effects. The results show few significant effects: states with a larger population actually experiment more, while states with a large immigrant population experiment less. More importantly, the influence of governor quality remains unchanged.

6 Conclusion

In this paper, we analyzed the role of policy makers’ reputational concerns for policy experimentation and reversal decisions. For that purpose, we constructed a novel data set on the landmark US Welfare Reform containing welfare policy innovations at the US state level between 1978 and 2010 and a wide range of political and socioeconomic controls. Predictions of a reputational model based on asymmetric information about policy makers and uncertainty about policy innovations are confirmed by the data. We find statistically and economically significant effects showing the importance of reputational concerns for policy experiments and reversals.

²⁰The set of demographics controls we include are those commonly used in other studies of policy innovations; see Walker (1969); Berry and Berry (1992) or Boehmke and Skinner (2012) for examples.

References

- Alesina, A. and Drazen, A. (1991). Why are stabilizations delayed? *American Economic Review*, 81(5):1170–1188.
- Allen, R. and Clark, J. (1981). State policy adoption and innovation: Lobbying and education. *State and Local Government Review*, 13(1):18–25.
- Bernecker, A. (2013). Divided we reform? evidence from us welfare policies. *Working Paper, University of Mannheim*.
- Berry, F. S. and Berry, W. D. (1990). State lottery adoptions as policy innovations: An event history analysis. *American Political Science Review*, 84(2):395–415.
- Berry, F. S. and Berry, W. D. (1992). Tax innovation in the states: Capitalizing on political opportunity. *American Journal of Political Science*, 36(3):715–742.
- Berry, F. S. and Berry, W. D. (2007). *Innovation and Diffusion Models in Policy Research*. In *Theories of the Policy Process*., volume second edition. Westview.
- Berry, W. D., Ringquist, E. J., Fording, R. C., and Hanson, R. L. (1998). Measuring citizen and government ideology in the American States, 1960-93. *American Journal of Political Science*, 42(1):327–348.
- Besley, T. and Case, A. (1995a). Does electoral accountability affect economic policy choices? Evidence from gubernatorial term limits. *Quarterly Journal of Economics*, 110(3):769–798.
- Besley, T. and Case, A. (1995b). Incumbent behavior: Vote-seeking, tax-setting, and yardstick competition. *American Economic Review*, 85(1):25–45.
- Besley, T. and Case, A. (2003). Political institutions and policy choices: Evidence from the United States. *Journal of Economic Literature*, 41(1):pp. 7–73.
- Blank, R. M. (2002). Evaluating welfare reform in the United States. *Journal of Economic Literature*, 40:1105–66.
- Boehmke, F. and Skinner, P. (2012). The determinants of state policy innovativeness. *working paper, University of Iowa*.

- Center for Economic and Policy Research (2012). *Current Population Survey ORG Uniform Extracts*, volume Version 1.7. Washington, D.C.
- Cnudde, C. F. and McCrone, D. J. (1969). Party competition and welfare policies in the American States. *American Political Science Review*, 63:858–66.
- Coate, S. and Morris, S. (1999). Policy persistence. *American Economic Review*, 89(5):pp. 1327–1336.
- Council of State Governments (2012). *Book of the States*. Council of State Governments,, Washington D.C. also earlier years.
- Crouse, G. (1999). Implementation of major changes to welfare policies, 1992-1998. http://aspe.hhs.gov/hsp/waiver-policies99/policy_cea.htm.
- Cukierman, A. and Tommasi, M. (1998). When does it take a Nixon to go to China? *American Economic Review*, 88(1):180–197.
- Fernandez, R. and Rodrik, D. (1991). Resistance to reform: Status quo bias in the presence of individual- specific uncertainty. *American Economic Review*, 81(5):1146–1155.
- Francis, R. (1998). Predictions, patterns, and policymaking: A regional study of devolution. *Publius: The Journal of Federalism*, 28:143–160.
- Fu, Q. and Li, M. (2010). Policy-making with reputation concerns. Technical report, working paper, National University of Singapore.
- Gagliarducci, S. and Nannicini, T. (2013). Do better paid politicians perform better? disentangling incentives from selection. *Journal of the European Economic Association*. forthcoming.
- Galasso, V. and Nannicini, T. (2011). Competing on good politicians. *American Political Science Review*, 105(01):79–99.
- Grogger, J. and Karoly, L. A. (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., J. Camasso, M., and Jagannathan, R. (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.
- 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 Mukand, S. W. (2003). Redistributive promises and the adoption of economic reform. *American Economic Review*, 93(1):256–264.
- Kaplan, T. (2000). Wisconsin works. In Liebschutz, S. F., editor, *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 Poutvaara, P. (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>.
- Lieberman, R. C. and Shaw, G. M. (2000). Looking inward, looking outward: The politics of state welfare innovation under devolution. *Political Research Quarterly*, 53(2):215–240.

- List, J. A. and Sturm, D. M. (2006). How elections matter: Theory and evidence from environmental policy. *The Quarterly Journal of Economics*, 121(4):1249–1281.
- Majumdar, S. and Mukand, S. W. (2004). Policy gambles. *American Economic Review*, 94(4):1207–1222.
- Mead, L. M. (2004). *Government Matters. Welfare Reform in Wisconsin*. Princeton University Press.
- Moen, E. R. and Riis, C. (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 Danziger, S., editor, *Economic Conditions and Welfare Reform*. W.E. Upjohn Institute.
- Moffitt, R. A. (2002). Welfare Benefits File. <http://www.econ.jhu.edu/People/Moffitt>.
- Preuhs, R. R. (2006). The conditional effects of minority descriptive representation: Black legislators and policy influence in the American States. *Journal of Politics*, 68:585–99.
- 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.
- Soss, J., Schram, S. F., Vartanian, T. P., and O’Brien, E. (2001). Setting the terms of relief: Explaining state policy choices in the devolution revolution. *American Journal of Political Science*, 45(2):378–395.
- Tweedie, J. (2000). From D.C. to Little Rock: Welfare reform at mid-term. *Publius*, 30:69–97.
- U. S. Department of Homeland Security (2011). *Yearbook of Immigration Statistics*. Office of Immigration Statistics, Washington, D.C. also earlier years.
- United States Census Bureau (2011). *Statistical Abstract of the United States: 2012* (131st edition).

Urban Institute (2012). Welfare Rules Database. Washington, D.C.

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.

Winston, P. (2002). *Welfare Policymaking in the States: The Devil in Devolution*. Georgetown University Press.

A Data Appendix

A.1 Welfare Waivers and Policy Rules

To obtain measures of 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 information in the WRD is often more detailed and up-to-date than the official plans states submit periodically to the federal government (providing an overview of states' choices under the block grant). Our main analysis focuses on twelve rules in the four policy areas that were at the center of the public and political debate surrounding welfare reform.

Family Caps. Under AFDC, a family's benefit level depended upon its size, so if a recipient had a baby the grant amount rose. Family caps allowed states to eliminate or reduce the increase in benefits when an additional child was born. All family cap waivers except New Jersey's exempt children conceived as a result of rape or incest from the family cap. Several states, such as Wisconsin, Massachusetts and Illinois, specify that a child born or conceived after a family no longer receives AFDC can be denied benefits if the family returns to AFDC.

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 put into exemption categories (because of age, illness or having a small child). Under TANF, states require most adult heads of an assistance unit to perform some type of work-related activity after a given period. Work programs vary widely from state to state including 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; hours requirements vary from a mere good faith effort to find a job to fulltime 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 work requirement time limit if a parent fails to work at least 20 hours per week in a real job after a certain number of assistance months.

Sanctions. The Job Opportunities and Basic Skills Training Program (JOBS), part of the 1988 Family Support Act, provides education, training and work experience activities to AFDC recipients. Under TANF, states require most adult heads of an assistance unit to perform some type of work-related activity after a given period. If a benefit unit does not comply with these requirements, states can impose different sanctions. The first rule characterizes the initial sanction if the benefit unit does not comply with the work requirements. The initial sanction varies from reduction of 25% or less to a suspension of the full family benefit. The other three rules apply to the severity of the worst sanction 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 which ranges 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 we code 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 family. The third rule defines the duration of the lifetime limit ranging from no time limit to only 24 months. The final rule defines whether the state allows for any extensions to the lifetime limit or not.

A.2 Politics and Ideology Measures

Information on US governors was collected from the website of the National Governor's Association and merged to data kindly provided by David J. Andersen from the Eagleton Institute of Politics at Rutgers University. The quality of a governor is calculated as years in political office, more precisely as the number of years between taking 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. The low quality dummy is equal to one for all governors with this quality measure being below the median (and zero otherwise). The data for the party composition of the state legislature, the party of the governor and indicators for a divided government come from Klarner (2003). Polarization 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. Information on women legislators and women governors was collected from the website of the Center for American Women and Politics. Data for women legislators are available biannually from 1975 to 1983 and annually thereafter. The data for African American legislators for 1984 to 1993 is from Preuhs (2006) and for 1994 to 2005 from the website of the National Conference on State Legislatures. Data on term limits and vote margins for governors from 1960 to 2000 are taken from List and Sturm (2006) and updated to 2011 using the Book of the States and the website <http://uselectionatlas.org/RESULTS/>. Data for term limits of state legislators are from the National Conference on State Legislatures. We use measures of voter and government ideology calculated by Berry et al. (1998) from 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. 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.

A.3 State Demographics

TANF caseloads for both fiscal year and calendar year are taken from Moffitt (2002) for 1960 to 1998 and updated to 2010 using the Statistical Abstract. Average monthly payments for TANF recipients and families (by fiscal year) are also from Moffitt (2002) for 1960-1997 and updated to 2010 using the Annual Statistical Supplement to the Social Security Administration. Job entry, job retention and earnings gain data for 1997-2010 are taken from the High Performance Bonus data (table 5.4) reported in the TANF-Annual Reports to Congress. The out-of-wedlock birth rate (births to unmarried women per 1,000 unmarried women between 15 and 44 years of age) is available for 1992 to 2003 from table 8.3 in the TANF Annual Reports to Congress and updated to 2010 from the Vital Statistics. Resident population and resident population by age (under 5, age 5 to 17, over 18, over 65) and the population by race (Whites, Blacks and Hispanic) is taken from the Statistical Abstract. Personal income per capita, transfer receipts, wage income and total employment is taken from the website of the Bureau of Economic Analysis. The unemployment rate for 1960 to 1998 is from Moffitt (2002) and updated for 1998 to 2010 from the website of the Bureau of Labor Statistics. Indicators of income inequality (standard deviation, 90th/10th percentile ratio and Gini coefficient of total family income, personal income for men and women as well as wage and salary payments for men and women) are calculated from the March Current Population Survey. All income variables are deflated by the urban consumer price index with base year 2002.

Table 1: 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 2: Summary Statistics

Variable	Waiver Period (1978-1996)			TANF Period (1996-2010)		
	N	Mean	Std. Dev.	N	Mean	Std. Dev.
# Policy Experiments (Waiver Applications)	969	.2012384	.5765372			
Indicator for Waiver Experimentation	969	.1496388	.3569012			
# Post 1996 Experiments				714	.8137255	1.550562
Indicator for post 1996 Experimentation				714	.3431373	.4750897
# Post 1996 Reversals				714	.1918768	.4950143
Indicator for post 1996 Reversals				714	.1526611	.3599126
Governor Quality (Years in Political Office)	903	11.43189	9.096402	664	12.04518	9.759293
Low Governor Quality	903	.4983389	.5002743	664	.4623494	.4989563
Governor Age	935	55.15187	9.585323	694	59.00576	8.527857
Governor Lame Duck	950	.2378947	.4260186	650	.3169231	.4656355
Governor Past Vote Margin	913	8.367755	7.063581	674	12.45428	11.17812
Governor Party	950	.5871744	.4886193	700	.4364286	.4923173
Governor Party Ideology (Berry et al. 1998)	950	-.0185552	.2790226	700	.0593503	.3188724
Citizens Ideology (Berry et al. 1998)	950	46.96159	15.2925	600	51.29706	15.99161
Democratic Presidential Vote Share	850	41.53648	7.089873	703	47.08796	9.035134
Divided Government	931	.5488722	.4978732	679	.5478645	.4980706
Democratic Seat Share Upper House	931	.601858	.1859281	686	.5116806	.1621021
Democratic Seat Share Lower House	931	.5934898	.1800033	686	.5193672	.1535433
Polarization Upper House	931	.1703237	.1261511	686	.131253	.0957144
Polarization Lower House	931	.1638331	.1195023	686	.1252652	.0907574
% AFDC/TANF Recipients in Population (Caseload)	950	.0404008	.0144649	713	.0187619	.0139151
Unemployment Rate	950	6.465789	2.110039	711	5.289733	1.88728
Per Capita Income (/1000)	969	15.71918	5.310862	714	32.96318	7.456562
Population (/1000)	969	4.782394	5.19861	714	5.715069	6.360649
% Black Population	950	9.399228	9.157036	709	10.85607	10.71709
% Population 65+	950	11.94104	2.27207	709	12.72292	1.789727
% Immigrant Population	951	1.761939	2.215588	714	2.260454	1.718433
Gini Coefficient (Family Income)	950	.3575397	.0246252	400	.4014018	.0220994
Av. # Experiments in Adjacent States	950	.2219035	.4076696			
Av. # Experiments in States with similar Population	969	.2034744	.3502401			

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, see the data appendix for details. The low quality governor indicator is equal to one if the number of years of political experience is below the median value. Governor Lame 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%. Adjacent states are geographically adjacent states, states with similar population refers to states with similar population size (where all states are divided into ten bands of similar population size).

Table 3: Electoral Incentives and Policy Experimentation during the Waiver Period (1978-1996)

	<u>All Governors</u>			<u>Differential Effect for Lameducks</u>		
	(1)	(2)	(3)	(4)	(5)	(6)
Quality * Caseload in State	0.635*** (0.189)	0.491* (0.245)	0.511** (0.249)	0.733*** (0.206)	0.583** (0.267)	0.586** (0.283)
Quality	-0.0207*** (0.00710)	-0.0170* (0.00892)	-0.0181* (0.00918)	-0.0243*** (0.00770)	-0.0192** (0.00934)	-0.0195* (0.0104)
Caseload in State	-1.313 (1.874)	-5.647 (5.156)	-5.817 (5.153)	-1.416 (1.855)	-5.438 (5.188)	-5.402 (5.186)
Governor Age			0.000871 (0.00156)			0.000508 (0.00176)
Lameduck * Quality * Caseload in State				-0.436** (0.217)	-0.399 (0.258)	-0.316 (0.349)
Lameduck * Quality				0.0163** (0.00791)	0.00992 (0.00861)	0.00537 (0.0132)
Lameduck * Caseload in State				-0.550 (1.014)	3.235* (1.916)	1.326 (4.038)
Lameduck * Governor Age						0.00166 (0.00291)
Year Fixed Effects	YES	YES	YES	YES	YES	YES
State Fixed Effects	NO	YES	YES	NO	YES	YES
Observations	846	846	846	846	846	846
R-squared	0.195	0.284	0.284	0.200	0.287	0.287

Notes: The dependent variable in all specifications is the number of policy experiments (waiver applications) in a given state and year. Specifications (1) and (4) include year fixed effects, all other specifications additionally include state fixed effects. The quality of a governor is measured by years of political experience, see the data appendix for details. Caseload refers to the percentage of AFDC/TANF recipients in the population, lagged by one year. The lame duck dummy is equal to one if the governor cannot be reelected. 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 Experimentation in the Post-TANF Period (1996-2010)

	<u>All Governors</u>			<u>Differential Effect for Lameducks</u>		
	(1)	(2)	(3)	(4)	(5)	(6)
Quality * Caseload in State	0.799** (0.375)	1.226** (0.580)	1.228** (0.582)	0.739 (0.599)	1.550* (0.875)	1.671* (0.911)
Quality	-0.0124* (0.00684)	-0.0112 (0.0105)	-0.0124 (0.0111)	-0.0117 (0.00887)	-0.0173 (0.0132)	-0.0223 (0.0140)
Caseload in State	-13.69** (5.605)	-16.72 (14.63)	-17.31 (14.94)	-14.14** (6.926)	-15.39 (15.63)	-17.96 (16.39)
Governor Age			0.00352 (0.00628)			0.00614 (0.00684)
Lameduck * Quality * Caseload in State				0.208 (0.638)	-0.344 (0.824)	-0.620 (1.036)
Lameduck * Quality				-0.00276 (0.00862)	0.0129 (0.0123)	0.0214 (0.0205)
Lameduck * Caseload in State				-1.112 (8.851)	-6.011 (10.22)	-0.631 (15.98)
Lameduck * Governor Age						-0.00297 (0.00518)
Year Fixed Effects	YES	YES	YES	YES	YES	YES
State Fixed Effects	NO	YES	YES	NO	YES	YES
Observations	666	666	666	621	621	621
R-squared	0.383	0.406	0.407	0.379	0.403	0.403

Notes: The dependent variable in all specifications is the number of policy experiments (policy rule changes) in a given state and year. Specifications (1) and (4) include year fixed effects, all other specifications additionally include state fixed effects. The quality of a governor is measured by years of political experience, see the data appendix for details. Caseload refers to the percentage of AFDC/TANF recipients in the population, lagged by one year. The lame duck dummy is equal to one if the governor cannot be reelected. Standard errors clustered at the state level are shown in parentheses. *** p<0.01, ** p<0.05 and *p<0.1.

Table 5: Initial Reputation and Policy Experimentation

	<u>All Governors</u>			<u>Differential Effect for Lameducks</u>		
	(1)	(2)	(3)	(4)	(5)	(6)
Past Governor Vote Margin	-0.00680** (0.00305)	-0.00477* (0.00276)	-0.00447 (0.00277)	-0.00869*** (0.00307)	-0.00759** (0.00332)	-0.00743** (0.00366)
Governor Age			0.000652 (0.00145)			0.000293 (0.00144)
Lameduck * Past Governor Vote Margin				0.00491* (0.00263)	0.00744** (0.00281)	0.00816* (0.00481)
Lameduck * Governor Age						-0.000212 (0.00121)
Year Fixed Effects	YES	YES	YES	YES	YES	YES
State Fixed Effects	NO	YES	YES	NO	YES	YES
Observations	864	864	849	864	864	849
R-squared	0.170	0.278	0.277	0.172	0.281	0.281

Notes: The dependent variable in all specifications is the number of policy experiments (waiver applications) in a given state and year. Specifications (1) and (4) include year fixed effects, all other specifications additionally include state fixed effects. The past governor vote margin measures the winner's distance in votes to the runner-up in the last gubernatorial election. The lame duck dummy is equal to one if the governor cannot be reelected. Standard errors clustered at the state level are shown in parentheses. *** p<0.01, ** p<0.05 and *p<0.1.

Table 6: Electoral Incentives and Policy Reversals in the Post-TANF Period (1996-2010)

	<u>All Governors</u>			<u>Differential Effect for Lameducks</u>		
	(1)	(2)	(3)	(4)	(5)	(6)
Low Quality * Caseload in State	-4.532** (1.851)	-4.195* (2.441)	-4.277* (2.435)	-4.204* (2.316)	-3.754 (2.704)	-4.740* (2.676)
Low Quality	0.0949** (0.0442)	0.0680 (0.0452)	0.0852* (0.0459)	0.0899 (0.0607)	0.0560 (0.0552)	0.0945 (0.0603)
Caseload in State	1.417 (1.151)	-0.212 (3.023)	-0.611 (2.954)	0.768 (1.651)	-0.469 (3.370)	-0.108 (3.470)
Governor Age			0.00268 (0.00199)			0.00268 (0.00210)
Lameduck * Low Quality * Caseload in State				-0.359 (3.624)	-0.863 (3.806)	1.309 (4.351)
Lameduck * Low Quality				0.0116 (0.0755)	0.0221 (0.0734)	-0.0309 (0.104)
Lameduck * Caseload in State				1.220 (1.794)	0.709 (1.981)	-1.183 (2.819)
Lameduck * Governor Age						0.000649 (0.00137)
Year Fixed Effects	YES	YES	YES	YES	YES	YES
State Fixed Effects	NO	YES	YES	NO	YES	YES
Observations	666	666	666	621	621	621
R-squared	0.155	0.216	0.218	0.155	0.224	0.226

Notes : The dependent variable in all specifications is an indicator whether the state has reversed any policy rule in a given year or not. Specifications (1) and (4) include year fixed effects, all other specifications additionally include state fixed effects. The low quality governor indicator is equal to one if the number of years of political experience is below the median value. Caseload refers to the percentage of AFDC/TANF recipients in the population, lagged by one year. The lame duck dummy is equal to one if the governor cannot be reelected. Standard errors clustered at the state level are shown in parentheses. *** p<0.01, ** p<0.05 and *p<0.1.

Table 7: The Role of Ideology for Policy Experimentation

	(1)	(2)	(3)	(4)	(5)	(6)
Caseload in State * Republican Governor	-0.345 (3.850)					
Caseload in State * Democratic Governor	0.558 (4.107)					
Quality * Caseload in State		0.519** (0.254)	0.529** (0.257)	0.510** (0.249)	0.568** (0.269)	0.583** (0.287)
Quality		-0.0183* (0.00935)	-0.0186* (0.00949)	-0.0181* (0.00920)	-0.0196* (0.0101)	-0.0200* (0.0107)
Caseload in State		-5.831 (5.149)	-5.801 (5.124)	-5.801 (5.195)	-7.673 (5.635)	-7.277 (5.701)
Governor Age		0.000963 (0.00155)	0.00104 (0.00159)	0.000867 (0.00156)	0.000304 (0.00194)	0.000249 (0.00198)
Governor Party Dummy		-0.0155 (0.0380)				0.192 (0.135)
Governor Party Ideology (Berry et al. 1998)			0.0609 (0.0770)			0.381 (0.276)
Citizens Ideology (Berry et al. 1998)				-0.000213 (0.00277)		0.000820 (0.00381)
Democratic Presidential Vote Share					0.00746 (0.00605)	0.00634 (0.00605)
Year Fixed Effects	YES	YES	YES	YES	YES	YES
State Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	880	846	846	846	758	758
R-squared	0.274	0.284	0.284	0.284	0.290	0.291

Notes : The dependent variable in all specifications is the number of policy experiments (waiver applications) in a given state and year. All specifications include year and state fixed effects. The quality of a governor is measured by years of political experience, see the data appendix for details. Caseload refers to the percentage of AFDC/TANF recipients in the population, lagged by one year. 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 last presidential election. Standard errors clustered at the state level are shown in parentheses. *** p<0.01, ** p<0.05 and *p<0.1.

Table 8: Other Political Factors and Policy Experimentation

	(1)	(2)	(3)	(4)
Quality * Caseload in State	0.573** (0.255)	0.503** (0.249)	0.523** (0.244)	0.565** (0.255)
Quality	-0.0201** (0.00947)	-0.0180* (0.00915)	-0.0191** (0.00893)	-0.0202** (0.00937)
Caseload in State	-6.456 (5.260)	-5.536 (5.193)	-5.860 (5.151)	-6.259 (5.347)
Governor Age	0.00103 (0.00177)	0.000924 (0.00155)	0.000868 (0.00157)	0.00103 (0.00175)
Divided Government	0.0844** (0.0374)			0.0695* (0.0407)
Dem. Seat Share Upper House		-0.270 (0.273)		-0.161 (0.265)
Dem. Seat Share Lower House		-0.0356 (0.316)		-0.00224 (0.321)
Polarization Upper House			-0.361 (0.350)	-0.220 (0.345)
Polarization Lower House			-0.160 (0.309)	-0.0839 (0.319)
Year Fixed Effects	YES	YES	YES	YES
State Fixed Effects	YES	YES	YES	YES
Observations	835	846	846	835
R-squared	0.288	0.285	0.286	0.289

Notes : The dependent variable in all specifications is the number of policy experiments (waiver applications) in a given state and year. All specifications include state and year fixed effects. The quality of a governor is measured by years of political experience, see the data appendix for details. Caseload refers to the percentage of AFDC/TANF recipients in the population, lagged by one year. 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%. Standard errors clustered at the state level are shown in parentheses. *** p<0.01, ** p<0.05 and *p<0.1.

Table 9: Spillovers between States and Policy Experimentation

	(1)	(2)	(3)	(4)	(5)	(6)
Quality * Caseload in State	0.546** (0.262)	0.561** (0.269)	0.626** (0.300)	0.528** (0.251)	0.534** (0.255)	0.577* (0.293)
Quality	-0.0201** (0.00991)	-0.0193* (0.00999)	-0.0222* (0.0111)	-0.0197** (0.00950)	-0.0186* (0.00947)	-0.0205* (0.0106)
Caseload in State	-6.025 (5.443)	-6.717 (5.344)	-5.650 (5.939)	-6.965 (5.227)	-8.101 (4.950)	-6.926 (5.444)
Governor Age	0.00135 (0.00179)	0.000731 (0.00206)	0.00230 (0.00203)	0.00117 (0.00176)	0.000325 (0.00198)	0.00184 (0.00193)
Geographic Neighbors' Experiments (t-1)	-0.0978 (0.0840)	-0.124 (0.0846)	-0.124 (0.0880)			
Geographic Neighbors' Experiments (t-2)		-0.123 (0.0877)	-0.117 (0.0936)			
Geographic Neighbors' Experiments (t-3)			0.0554 (0.106)			
Pop. Size Neighbors' Experiments (t-1)				0.247 (0.204)	0.253 (0.208)	0.266 (0.199)
Pop. Size Neighbors' Experiments (t-2)					0.0667 (0.158)	0.0779 (0.161)
Pop. Size Neighbors' Experiments (t-3)						-0.0397 (0.159)
Year Fixed Effects	YES	YES	YES	YES	YES	YES
State Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	802	758	713	802	758	713
R-squared	0.284	0.292	0.296	0.292	0.299	0.303

Notes: The dependent variable in all specifications is the number of policy experiments (waiver applications) in a given state and year. All specifications include year and state fixed effects. The quality of a governor is measured by years of political experience, see the data appendix for details. Caseload refers to the percentage of AFDC/TANF recipients in the population, lagged by one year. Geographic Neighbors' Experiments refers to the average value of the dependent variable for adjacent states (geographically neighboring states). Pop. Size Neighbors' Experiments refers to the average value of the dependent variable for states of similar population size (where all states are divided into ten bands of similar population size). Standard errors clustered at the state level are shown in parentheses. *** p<0.01, ** p<0.05 and *p<0.1.

Table 10: Socioeconomic Characteristics and Policy Experimentation

	(1)	(2)	(3)
Quality * Caseload in State	0.512** (0.245)	0.519** (0.247)	0.455* (0.251)
Quality	-0.0180** (0.00878)	-0.0179** (0.00886)	-0.0159* (0.00889)
Caseload in State	-8.558 (5.167)	-9.213* (5.151)	-9.471* (5.334)
Governor Age	-0.000149 (0.00159)	-0.000119 (0.00156)	-0.000365 (0.00147)
Unemployment Rate (/1000)	0.305 (12.52)	0.333 (12.46)	-2.295 (11.66)
Per Capita Income (/1000)	-0.00311 (0.0185)	-0.0119 (0.0195)	-0.00326 (0.0202)
Population (/1000)	0.106*** (0.0318)	0.122*** (0.0322)	0.167*** (0.0390)
% Population Black		0.0480 (0.0604)	0.0577 (0.0573)
% Population 65+		0.0313 (0.0425)	0.0318 (0.0442)
% Immigrant Population			-0.0504*** (0.0128)
Gini Coefficient (Family Income)			-0.252 (1.056)
Year Fixed Effects	YES	YES	YES
State Fixed Effects	YES	YES	YES
Observations	846	846	846
R-squared	0.293	0.294	0.301

Notes : The dependent variable in all specifications is the number of policy experiments (waiver applications) in a given state and year. All specifications include state and year fixed effects. The quality of a governor is measured by years of political experience, see the data appendix for details. Caseload refers to the percentage of AFDC/TANF recipients in the population, lagged by one year. Robust standard errors clustered at the state level are shown in brackets. *** p<0.01, ** p<0.05 and *p<0.1.

Figure 1: AFDC Caseload per Capita 1970-2000

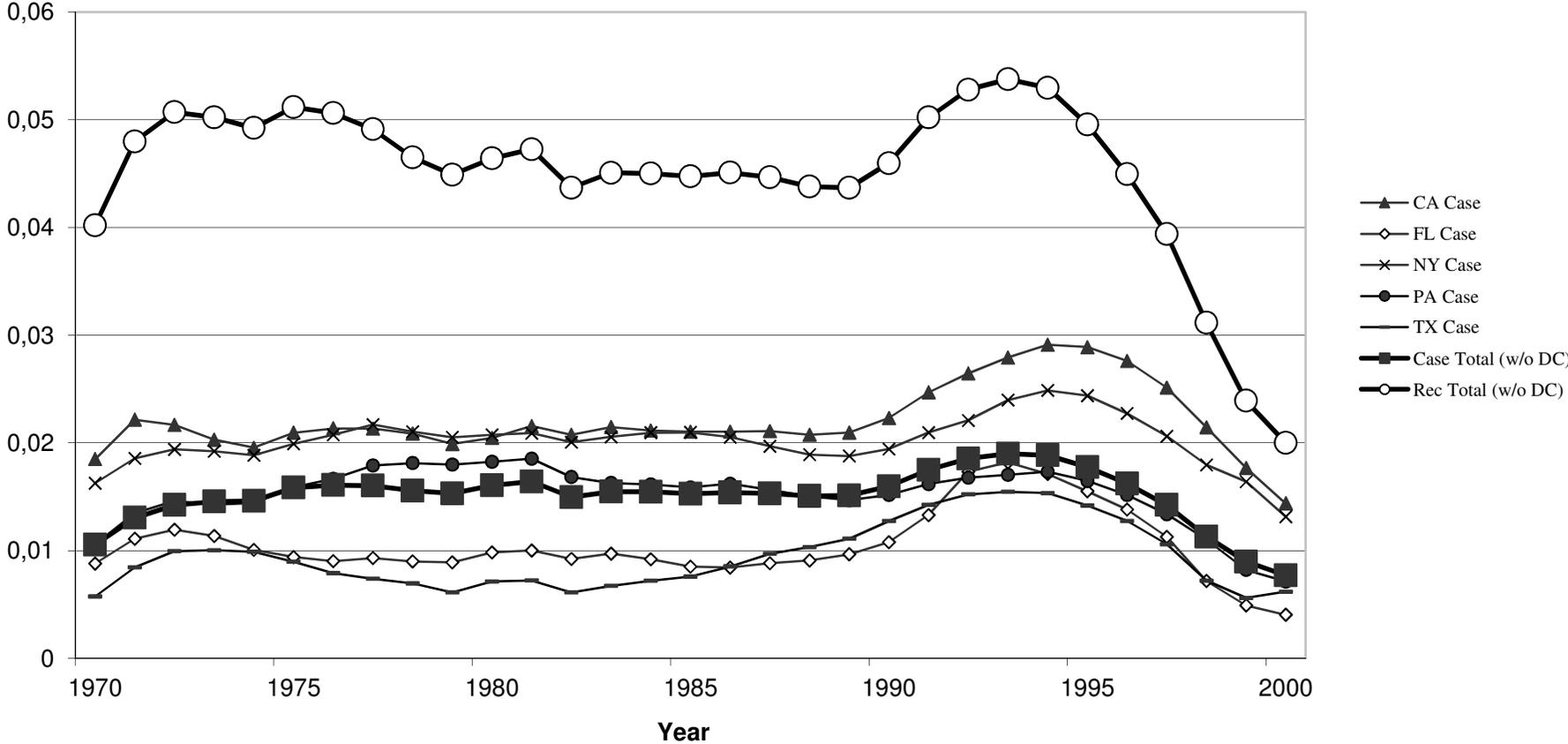
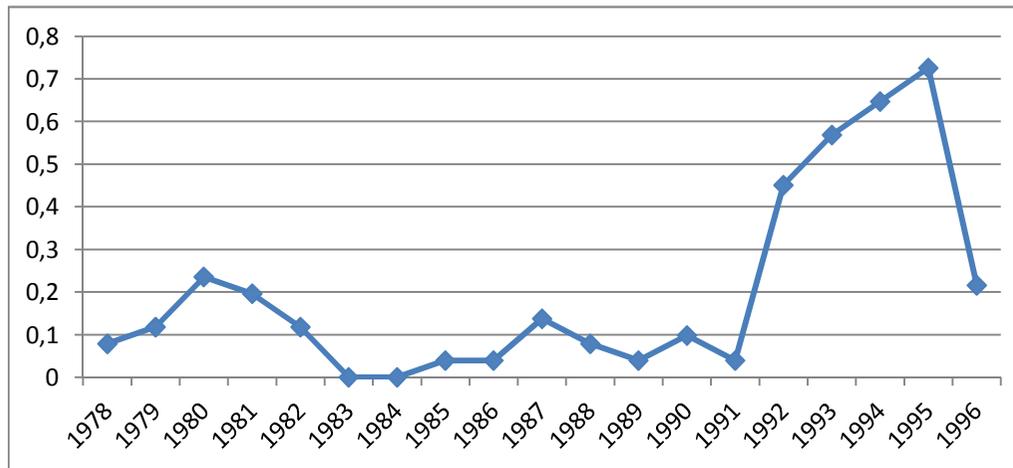
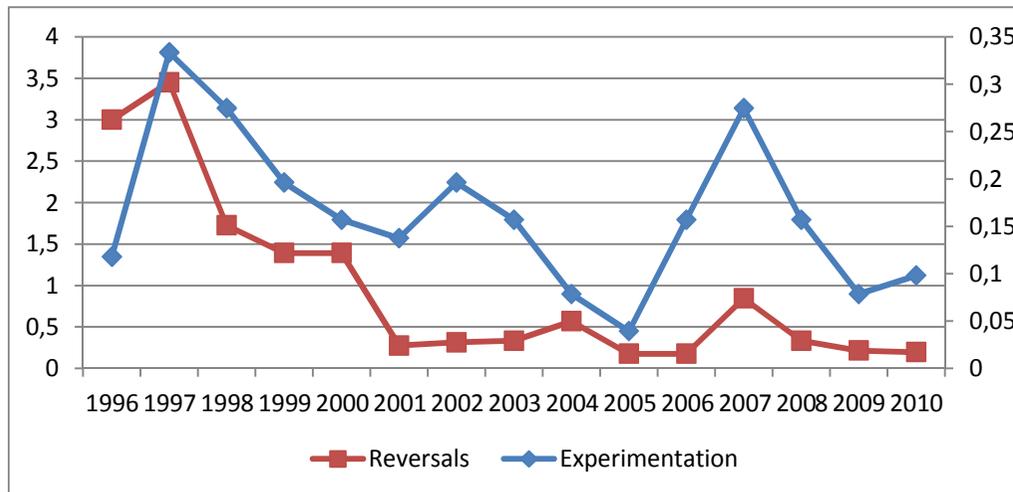


Figure 2: Evolution of Welfare Waivers (1978-1996)



Source: Authors' calculation

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



Source: Authors' calculation

Table A1: Coding of Policy Experimentation and Reversals (1996-2010)

Policy Rule	Description of Rule	Experimentation = 1	Reversal = 1
<u>Family Cap (1)</u>	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
<u>Work Requirements (3)</u>			
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.
<u>Time Limits (4)</u>			
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 intermittent time limit.	State reduces benefits more after unit hits intermittent time limit.	State reduces benefits less after unit hits intermittent time limit.
Time Limit Extensions	Whether the state offers any type of time limit extension.	State abolishes all time limit extensions.	State reintroduces at least one time limit extension.
<u>Sanctions (4)</u>			
Severity of Worst Sanction	How much benefits are reduced under worst sanction for non-compliance with work requirements.	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.

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

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