Dynamic Estimation of Policy Effects: An Application to Welfare Time Limits

Marc K. Chan *
Peking University

October 31, 2010

Abstract

The treatment effect of a program may often depend on the participant's past actions which is endogenous. We extend the standard linear dynamic panel data model and present a GMM estimator to solve this issue. The model is applied to Florida's Family Transition Program (FTP), a randomized experiment on welfare time limits. We find that the time limit caused the welfare receipt rate to drop by 6.4 percentage points for each additional year of past welfare receipt. This effect constituted around 40 percent of the overall behavioral effect of the time limit.

JEL CLASSIFICATION: C23, I3

Keywords: Treatment Effect, Dynamic Panel Data, GMM, Welfare, Time Limit

*This paper is loosely based on a chapter of my dissertation. I am grateful to Robert Moffitt for various discussions on this topic. I would like to thank Tony Burton, Peter Gottschalk, Jeffrey Grogger and Francesca Mazzolari for their comments. I would also like to thank Shiyu Li, Ying Zeng and Yue Zhang for research assistance. The data used in this paper are derived from data files made available to researchers by MDRC. The author remains solely responsible for how the data have been used or interpreted. Computer programs are available from the author. An earlier version was circulated as "Welfare-Banking when the Clock is Ticking: Another Look at Florida's Time-Limited Welfare Reform Experiment". Address: School of Economics, Peking University, Beijing, China. Email: mchanpku@gmail.com.

1 Introduction

In the experimental program evaluation literature, a pervasive strategy for impact analysis is to compare the mean outcomes between treatments and controls by months, quarters or years after random assignment.[1] When the treatment effect is suspected to vary by subgroups, mean outcomes between treatment and controls are usually separately compared in those subgroups of interest. These strategies do not apply when the analyst is interested in how the treatment effect depends on the treatment's past actions. For example, in a training program that helps trainees retain their jobs, the program should cause a larger increase in the employment probability this period among trainees who were employed last period. A similar argument can be applied to programs that attempt to eliminate a "dynamic welfare trap" (Plant (1984)). Yet another example is a training program that helps trainees with little work experience, so that the treatment effect may diminish when trainees accumulate work experience or when the job tenure increases. In these situations, conducting subgroup analysis or using interaction terms in ordinary least squares will produce inconsistent estimates, as past actions are endogenous and are likely to be correlated with unobserved characteristics.

In this paper, we extend the standard linear dynamic panel data model and present a generalized method of moments (GMM) estimator to solve this issue. We apply the estimator to a randomized experiment in Florida from 1994 to 1999 called the "Family Transition Program" (FTP) involving welfare time limits. A welfare time limit is a cap on the total amount of time that a family is allowed to participate in welfare. The family can choose when to participate, but once the cap is reached, it becomes ineligible for welfare. The treatment group was either limited to 24 months of cash assistance within any 60-month period (which we call "a 24-month limit"), or to 36 months of cash assistance within any 72-month period (i.e. which we call "a 36-month limit"). We find that the treatment effect increases substantially when the treatment group accumulates more periods of welfare receipt. Our model uncovers around 40 percent of the total treatment effect of the time limit not captured by existing reduced form models of time limits.

It is useful to relate the endogeneity issue with existing literature on standard dynamic models of individual behavior where the lagged dependent variable is present. In large N, small T settings, simple panel data estimators yield inconsistent estimates of state dependence (the effect of $y_{i,t-1}$ on y_{it}) in the presence of unobserved heterogeneity (e.g. Heckman (1978), Heckman (1981c), Heckman (1981a)). In a linear model with fixed or random effects, the endogeneity of the lagged dependent variable can be dealt with using further lags of the dependent variable as instruments (e.g. Anderson and Hsiao (1981a) and Anderson and Hsiao (1981b) for a simple instrumental variables (IV) estimator; Arellano and Bond (1991) for a GMM estimator). For dynamic probit and logit models, various estimators[2] are applied in applications such as labor force (Hyslop

[1] The treatment effect is therefore allowed to vary over time. In these studies, regression models, sometimes with weights, are usually employed to control for baseline factors. A recent application of this line of research methodology is Schochet, Burghardt, and McConnell (2008)'s study on the Job Corps program, a nationally representative experimental training program in the United States.

[2] A dynamic probit model with random effects can be estimated by simulation methods such as the smooth recursive conditioning (SRC) simulator (e.g. Borsch-Supan and Hajivassiliou (1993), Keane (1993), Keane (1994)) or Gaussian quadrature

(1999)) and welfare (Chay, Hoynes, and Hyslop (2001)) participation of women. These applications typically focus on the state dependence parameters but do not consider how certain policies may change them. In the experimental evaluation literature, a very limited amount of work have modeled dynamic behavior (e.g. Ham and LaLonde (1996) about training on employment and unemployment spells using duration models; Card and Hyslop (2005) about work subsidy on welfare participation using logit models). However, their focus is on disentangling the selection bias resulting from evaluating the experimental outcome of interest.[3] An exception is Card and Hyslop (2005), who interact treatment status with lagged welfare participation and accounts for endogeneity using a parametric random effects model.[4] Other works do not examine how a treatment can affect state dependence, not even mentioning incorporating specific "state variables" such as accumulated welfare participation or work experience.[5] In fact, our approach is not isolated but can be presented as a selection bias problem. To see this, consider a training program and an analyst defines two categories of work experience, high and low. The analyst would like to estimate the treatment effect of the high experience group.[6] If the treatment effect is a positive constant and unobserved heterogeneity is present, the high experience treatment group will have a lower unobserved effect on average than the high experience control group, so a simple subgroup estimator is downward biased. Similarly, the subgroup estimator for the low experience group is upward biased. Therefore, the analyst will erroneously conclude that the treatment effect decreases with experience.

We apply our GMM estimator to estimate the effects of time limits on welfare participation. Time limits were crucial in the welfare reform of the United States. The 1996 Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) stipulates that a female-headed family which has received federally-funded welfare assistance for a total of five years is ineligible for federal cash aid. As of 2006, 97 percent of the welfare caseloads in the United States were subject to a welfare time limit (Mazzolari and

(Butler and Moffitt (1982)). For dynamic logit models, conditional logit estimators for fixed effects are available (Cox (1958), Chamberlain (1985)).

[3]Ham and LaLonde (1996) argues that selection bias arises under an experimental setting if the outcome of interest (such as wage) is unobservable for part of the treatments and controls. In their paper that looks at a training program, they use a duration model with duration dependence and unobserved heterogeneity to correct for selection into fresh unemployment and employment spells after random assignment. Card and Hyslop (2005) looks at a work subsidy program that offers full-time work subsidy only if they establish eligibility by working a certain amount of time within the first 12 months after random assignment. They use a dynamic logit model with Gaussian quadrature and correct for selection into eligibles using a probit version of a hazard model. Both studies account for dynamic self-selection parametrically. A number of econometric techniques have been developed for static self-selection, e.g. Horowitz and Manski (2000) bounded treatment effect estimator and Lee (2009)'s bounded treatment effect estimator on wage in the Job Corps program.

[4]They find that for the treatment group, the conditional probability (given welfare status last period) of staying on welfare decreases. The conditional probability of staying off welfare increases, although the magnitude is 40 percent smaller. Since the focus was on selection bias, these results were reported in a table but not discussed in the text. No statistical test was conducted to test the equality of the two parameters.

[5]In duration models, "duration dependence" only allows the hazard rate to change exogenously with time. In parametric probit and logit models, state variables in the sense of accumulated past actions can be included. However, the initial conditions problem, if present (e.g. work experience is measured from the beginning of the life-cycle), may be hard to solve (Heckman (1981b)) compared to the linear framework than we present below. In addition, the interaction coefficient is harder to interpret in a nonlinear model (Ai and Norton (2003)). These state variables are usually dealt with in dynamic programming models (e.g. Keane and Wolpin (1997) on career decisions of men), but they do not focus on how the state variable changes the treatment effect of policies.

[6]We assume that the estimation is conducted on data some time after random assignment so that accumulated experience after random assignment is endogenous.

Ragusa (2009)). In Australia, Great Britain and New Zealand, time limits are currently under debate as a policy alternative to reduce the burgeoning welfare roll and to reduce "welfare dependency".[7]

There are three channels through which time limits can affect welfare participation. First, families that have reached the cap mechanically exhaust their benefits and become ineligible for welfare. Suppose a consumer with no saving decisions maximizes her expected lifetime utility. Then time limits will provide stronger incentives to conserve or bank benefits for families with younger children ("child effect" or the "GM effect" (Grogger and Michalopoulos (1999))). In addition, the incentive to bank benefits should increase when accumulated welfare receipt under a time limit increases ("stock effect" (Fang and Keane (2004))).[8] In both cases, time limits generate an option value for forward-looking families.

The literature on time limits has found sizable child effects (Grogger and Michalopoulos (2003), Grogger (2003), Grogger (2004)) and mechanical effects (Mazzolari (2007), Mazzolari and Ragusa (2009)). No reduced form models satisfactorily estimate the magnitude of the stock effect. For example, Fang and Keane (2004) use the Current Population Survey, a cross-sectional data. They proxy the stock by the length of time passed since the time limit was put into effect. Mazzolari (2007) and Mazzolari and Ragusa (2009) use the Survey of Income and Program Participation, a panel data. They circumvent the endogeneity problem by imputing the stock using pre-time-limit welfare participation rates that vary by observable family characteristics.[9] The stock is treated as a state variable in life-cycle dynamic programming models such as Swann (2005) and Keane and Wolpin (forthcoming), who simulate time limits using "ex-ante" data.[10] Although they report combined child and stock effects, their models are capable of generating stock effect estimates.

Our model finds a sizable stock effect. First, the child and stock effects combined caused welfare participation to drop by around 29 percent. In contrast, Grogger and Michalopoulos (2003), who models the child effect only, found the reduction to be 16 percent based on similar data. Second, the stock effect constituted 40 percent of the reduction. Third, the mechanical effect was much smaller than found in existing studies. In a counterfactual where we disallow the stock effect, the mechanical effect became considerably larger, suggesting a crowding-out phenomenon due to model misspecification. Fourth, the probability of welfare receipt dropped by an average of 6.4 percentage points for every additional year of welfare receipt under the time limit. In contrast, the probability of welfare receipt only dropped by 0.6 percentage points for every additional year of age of the youngest child in the family. In the terminology of intertemporal optimization, this suggests that the welfare-banking behavior of families was much more sensitive to the state variable

[7] The debate has been documented in the popular press, for example, Karvelas (2008), Hari (2008), and Rich (2003). In British Columbia, Canada, the time limit enjoyed very little popularity and was abandoned in 2004, three years after it was put in place (Bruce Wallace and Tim Richards (2008)).

[8] Alternatively, the incentive to bank benefits should increase when the months of remaining eligibility under a time limit decreases. In most cases, both are equivalent. We will alert the reader whenever they are different.

[9] In addition, their models do not estimate the stock and the child effect as separate parameters. Using proxies for the stock results in identification issues in some circumstances. For example, the Fang and Keane model will be unidentified if there is no variation in the implementation date of the time limit. In the Mazzolari model, a similar issue will occur if there is no variation in the type of the time limit, or if the type of the time limit is endogenous.

[10] Swann uses the Panel Study of Income Dynamics (PSID) from 1968 to 1992; Keane and Wolpin use the National Longitudinal Survey of Youth (NLSY) from 1979 to 1991.

(i.e. total time of welfare receipt) than the time horizon (i.e. age of the youngest child). Fifth, the incentive to bank benefits increases nonlinearly with the stock. Using a spline specification, we find that for families with less than 1 year of welfare eligibility remaining, one marginal year of welfare receipt under the time limit caused the probability of welfare receipt to drop by 17 to 20 percentage points.

This paper is organized as follows. Section 2 describes Florida's Family Transition Program and the data. Section 3 describes our empirical model of time limits and outlines the econometric issues involved. Section 4 develops a dynamic panel data model and the corresponding GMM estimator, and applies them to our time limit model. Section 5 reports the results. In that section, we compare the GMM model with OLS and simple panel data models and perform robustness check. We access the within-sample fit and extrapolation abilities of our model and compare them with the GM and Mazzolari models. We perform various counterfactual simulation exercises. We also estimate our model by decomposing the program group into 24- and 36-month limit subgroups. Section 6 concludes.

2 Florida's Family Transition Program

Florida's Family Transition Program (FTP) was the first welfare reform initiative in the United States to impose a time limit. It was a pilot project that operated from 1994 to 1999 in Escambia, a mid-sized county in the northwest of the panhandle region in Florida. From May 1994 through October 1996, welfare applicants who met FTP eligibility were randomly assigned to the Aid to Families with Dependent Children (AFDC) group or the FTP group. Being the control group, the AFDC group was subject to welfare rules that existed before FTP was implemented. Being the program group, the FTP group was subject to rules under the FTP program which was different from AFDC in the following key components:

1. Time limit: Approximately half of the FTP group were limited to 24 months of cash assistance within any 60-month period (which we call "a 24-month limit"). The other half, who were determined as more disadvantaged, were limited to 36 months of cash assistance within any 72-month period (i.e. which we call "a 36-month limit"). There was no time limit in the AFDC group;

2. Financial work incentives: FTP participants were given a more generous earnings disregard (the first 200 dollars and 50 percent of any remaining earnings). The AFDC group had a earnings disregard of the first 120 dollars and 33 percent of any remaining earnings;

3. Enhanced services: FTP participants received more intensive case management and enhanced employment, training and other support services;

4. Parental responsibility mandates: Under FTP, parents held more responsibility regarding children's development.

Since the FTP group differs from the AFDC group in several components, their separate effects are unidentifiable with a treatment status variable in the equation (Grogger and Michalopoulos (2003)). Static labor supply theory predicts that a more generous earnings disregard increases welfare participation.[11] Enhanced services may have an ambiguous effect on welfare participation, and parental responsibility mandates may reduce welfare participation.[12] The effects of the time limit can be identified in two ways. First, welfare participation should drop more for families with younger children due to the child effect, which was the empirical approach adopted by Grogger and Michalopoulos (2003). Second, the stock effect implies that welfare participation should drop more for families that have received welfare for more months, other things unchanged. [13] Notice that only the accumulated months on welfare matters, but not the pattern of welfare spells.

Basic demographic information at the time of random assignment (which we call "month 0") was collected via the Background Information Form (BIF). Administrative records on quarterly earnings and monthly AFDC/TANF and Food Stamps receipt were available from 24 months prior to the random assignment to 53 months after the random assignment. The data we obtained contains 2817 individuals who were randomly assigned between May 1994 and February 1995. This sample, called the "report sample", is the public use version of the data used in the original evaluation report (Bloom and et al. (2000)).[14] After excluding individuals with missing data in the BIF, we end up with 2570 individuals. Due to start-up issues and implementation problems, some applicants were not able to receive benefits within the first few months of random assignment (Bloom and et al. (2000)). To avoid the mixing-up of the behavioral effects and implementation constraints, our benchmark dataset starts from month 4 after random assignment. The benchmark dataset ends in month 23, one month prior to the families starting to exhaust their benefits. If our model has good predictive abilities, it should be able to accurately predict (by extrapolation) the welfare participation rate as well as the percentage of families exhausting benefits from month 24 onwards. Data from month 24 to 53 are used to assess the predictive abilities of the model only but not for estimation. Our panel has no attrition because complete administrative data are available. Another advantage is that our

[11] In addition, the amount of transfer income received should never fall; for example, see Bitler, Gelbach, and Hoynes (2006) for theoretical discussions and empirical findings on a welfare demonstration in Connecticut.

[12] Parents with schoolage children were required to ensure that their children attended school regularly and to speak with their children's teachers each grading period. Welfare applicants with pre-schoolage children must verify that their children have begun the necessary immunizations (Hendra (2002)). The effect of this mandate may vary by age but the sign is ambiguous.

[13] Grogger and Michalopoulos (2003) argues that the other three key policy elements should not have age-varying effects. We argue that their stock-varying effects should be at best minimal (if any). The most relevant scenario that we can imagine is the accumulation of human capital through increased work incentives. Consider the simplified case where a person works whenever she quits welfare, and does not work whenever she receives welfare. If human capital is an increasing function of total time worked, then a higher total time of welfare receipt would lead to less total time worked, less human capital and therefore a higher welfare receipt probability. Therefore, if human capital can be accumulated through work, the financial work incentive will offset the stock effect of a time limit and will make the magnitude of the stock effect biased downward. The human capital mechanism does not appear to hold well empirically, at least for low-income individuals. For example, using data from a policy experiment, Card and Hyslop (2005) shows that additional work efforts due to improved work incentives do not have lasting impact on wages.

[14] Although the full period of random assignment was from May 94 through Oct 96, subjects who were assigned beginning Mar 95 to Oct 95 were atypical due to programmatic reasons.

panel is longer and larger than typical panels where standard dynamic GMM methods (such as Arellano and Bond (1991)) are applied.

The summary statistics of major sample characteristics at the time of random assignment are reported in Table 1. The characteristics are similar to Grogger and Michalopoulos (2003) who use a larger sample.[15] The control and program groups are very similar (Bloom and et al. (2000) and Grogger and Michalopoulos (2003) show that the random assignment was carried out properly). The sample is disproportionately black and the education level is relatively low. The 36-month limit program group is more disadvantaged than the 24-month limit program group reflected by various demographic characteristics. Figure 1(a) shows the welfare participation rates of the control and program groups by month. The welfare participation rates for both groups peak in month 3, indicating that it takes time for applicants to receive benefits; start-up and implementation issues could also have delayed benefit receipt. From month 0 to 23, the welfare participation rate of the program group drops faster than the control group. A slight sudden drop of the welfare participation rate occurs in month 24, when around 2.5 percent of the program group exhaust benefits due to the 24-month limit. The gap between the control and program groups remain roughly constant until another drop occurs in month 36, when around 5 percent of the program group exhaust benefits due to the 36-month limit. From month 36 onwards, the welfare participation rate of the program group drops slower than the control group. Figure 1(b) compares the distribution of the total months of welfare receipt since random assignment between the control and program groups in month 24. The distributions are similar for both groups. More program group members have received welfare for a moderate number of months (between 7 and 18 months), and less program group members have received welfare for a small number of months (less than 7 months). The number of families having received welfare for a high number of months is comparable between both groups. This suggests that the FTP may have increased welfare participation but may have not done so uniformly across the distribution. The time limit may have prevented the program group members from receiving welfare for a large number of months. We will examine this formally in subsequent sections.

3 Empirical Model

We first consider the following dynamic panel data model of welfare participation:

$$\begin{aligned} y_{it} = \alpha_1 y_{i,t-1} + \alpha_2 y_{i,t-2} + \alpha_3 y_{i,t-3} + \beta_s d_i S_{it} + \beta_A d_i A_{it} + \\ \gamma_s S_{it} + \gamma_A A_{it} + \gamma_d d_i + X_{it}\gamma + \underbrace{\eta_i + \nu_{it}}_{\epsilon_{it}} \qquad t = 3, ..., T \end{aligned} \tag{1}$$

[15]The GM sample includes 4473 individuals who were randomly assigned between May 1994 and October 1996.

In this model, y_{it} is the welfare participation dummy of family i at time t, d_i is the program group dummy, $S_{it} \equiv \sum_{r=0}^{t-1} y_{ir}$ is the total months of welfare receipt since random assignment, A_{it} is the age of the youngest child in the family, X_{it} is a vector of regressors including mother's race, age and years of schooling, number of children in the family, a dummy for having the youngest child under 3 [16], an interaction term between this dummy and the program group dummy, a 36-month limit program subgroup dummy, year dummies, and a constant. The time series of each family is chosen such that no one exhausts their benefits. The error term ϵ_{it} consists of a permanent component (unobserved effect) η_i and a transitory component ν_{it}. In the benchmark model, the unobserved effect is not correlated with regressors in X nor to d and A. The robustness analysis relaxes this assumption. The unobserved effect is correlated with S and lagged y. In addition, S and lagged y are correlated with past values of the transitory error. Section 4 discusses the statistical properties of the model in further detail.

For the program group, S_{it} represents the stock under the time limit. The key parameters of interest are β_s and β_A, which represent the stock and the child effects respectively. We expect $\beta_s < 0$ and $\beta_A > 0$. The stock effect is linear in this model, but this assumption will be relaxed in a more complicated model in section 4. We model welfare participation behavior as dynamical following evidence from the U.S. (Chay and Hyslop (2000); Chay, Hoynes, and Hyslop (2001)), Canada (Hansen, Lofstrom, and Zhang (2006)) and Sweden (Andren (2007)). We choose 3 orders of autoregressive lags which is one order higher than the dynamic welfare model in Chay, Hoynes, and Hyslop (2001). We assume that $|\alpha_1|, |\alpha_2|, |\alpha_3| < 1$. The transitory errors are assumed to be serially uncorrelated (its validity will be tested by the Arellano and Bond (1991) m2 statistic). We assume that d, A, X are strictly exogenous, i.e. they are uncorrelated with past, concurrent, and future transitory errors (i.e. $E(X_{it}\nu_{is}) = 0$ for all t, s). In addition, we assume that d, A, X are uncorrelated with the unobserved effect.[17]

We consider the linear probability model for five reasons. First, direct comparison of the estimates can be made between this model and existing models, which are also linear. Second, the GMM estimator in the linear model is robust to potential initial condition problems related to S_{it}. Third, in nonlinear models such as probit (e.g. Hyslop (1999)), the marginal effects of the interaction terms are harder to interpret (Ai and Norton (2003)). Fourth, model specification tests such as the Sargan test of overidentifying restrictions and the m2 test (Arellano and Bond (1991)) are available for linear model. Fifth, the linear model provides semi-parametric identification.

Since there is more than one type of time limit, one might argue that welfare-banking behavior should depend on the months of remaining eligibility instead of the total time of welfare receipt under the time limit. In fact, both definitions yield identical estimates in the magnitude of the stock effect β_s. To see this, let $\bar{S}_i$ be the time limit length (e.g. 24 months or 36 months) of the program group, and let $\tilde{S}_{it} \equiv \bar{S}_i - S_{it}$

[16] For families with at least a child under 3, the control and the program groups differ in work-related activities. Grogger and Michalopoulos (2003) argue that the child effect cannot be identified in this case.

[17] In the robustness analysis, we allow A and X to be correlated with the unobserved effect.

be the months of remaining eligibility. Then the term $\beta_s d_i S_{it}$ can be rewritten as $-\beta_s d_i \tilde{S}_{it} + \beta_s d_i \bar{S}_i$. But the second term is a linear combination of the FTP group dummy and the 36-month limit dummy, so it is absorbed into both terms. Therefore using S_{it} or $\tilde{S}_{it}$ yield identical estimates of β_s in terms of magnitude, only that the sign is opposite.

To control for differences between the program and control groups as much as possible, we allow for S_{it} to have a direct effect on welfare receipt in the control group through the parameter γ_s. One reason for this inclusion is that some families in the control group erroneously believed that they were subject to a time limit, although many of them did not know what type of time limit they were subject to (Bloom and et al. (2000)). This is not surprising because the FTP experiment was carried out when the general decline of Florida's welfare caseload was unprecedented.[18] Starting October 1996, the control group of the FTP experiment was the only group in Florida that was not subject to a time limit. However, even if the time limit had an effect on the control group, we expect its magnitude to be less than the program group. Therefore, the estimated stock effect β_s should reflect the lower bound of the true stock effect. The second reason is that a mother may make welfare participation decisions conditional on the total time of welfare receipt since the life-cycle began, which we denote by $\tilde{\tilde{S}}_{it}$. In this case, S_{it} affects decisions directly even without the time limit. Fortunately, the model is free from the initial condition problem due to the unobservable $\tilde{\tilde{S}}_{it}$. To see this, note that we can express $\tilde{\tilde{S}}_{it}$ as the sum of S_i^0 and S_{it}, total time of welfare receipt before and after random assignment respectively. Then we have $\gamma_s \tilde{\tilde{S}}_{it} = \gamma_s S_i^0 + \gamma_s S_{it}$. But $\gamma_s S_i^0$ is time-invariant, so it is absorbed into the unobserved individual effect. Therefore, not knowing $\tilde{\tilde{S}}_{it}$ does not result in inconsistent estimates.

To end this section, it is useful to summarize the differences of our model from existing models of time limits. The most fundamental difference is that we model the stock effect β_s rigorously and estimate it as a separate parameter. The endogeneity of the stock makes the model considerably more complicated. In particular, our model becomes dynamic, and the resulting estimation strategy requires accounting for state dependence (Heckman (1981a)) and unobserved heterogeneity.

4 Estimation

A dynamic panel data model is necessary for the estimation of the stock effect because the total time of welfare receipt S is endogenous. In particular, it is positively correlated with the unobserved effect, and it is also a function of lagged dependent variables. If the autoregressive parameters $\alpha_1, \alpha_2, \alpha_3$ are zero, state dependence will contaminate the stock effect. If the unobserved individual effect is omitted, unobserved heterogeneity will contaminate the stock effect.

Estimation of the time limit model requires developing a variant of dynamic panel data models. To

[18] The welfare participation rate declined by 70 percent from 1994 to 1999 (Bloom and et al. (2000)).

see this, consider a simplified version of the empirical model with one order of autoregressive lag and $\beta_A, \gamma_s, \gamma_A, \gamma = 0$. The model can be rewritten as follows:

$$y_{it} = \alpha_1 y_{i,t-1} + \beta_s d_i \sum_{r=0}^{t-1} y_{ir} + \eta_i + \nu_{it} \tag{2}$$

Therefore, for the control group, the maximum autoregressive lag is of order 1; for the program group, the maximum autoregressive lag is of order t because every past period of welfare receipt under the time limit counts to the stock. In addition, the model is not stationary. Since the dynamics differ between the control and program groups, a new model is required. The following subsections develop the generic model and the corresponding GMM estimator, and discuss how a model of time limit can be applied.

4.1 Variable Autoregressive Error Components Model

Consider the following first order autoregressive model with an interaction with a dummy variable in the autoregressive component:

$$y_{it} = \alpha_1 y_{i,t-1} + \alpha_{1d} d_i y_{i,t-1} + X_{it}\gamma + \eta_i + \nu_{it} \qquad i = 1, ..., N; t = 1, ..., T \tag{3}$$

where $|\alpha_1|, |\alpha_{1d}| < 1$, d_i is a dummy[19], X_{it} is a K_x vector of strictly exogenous regressors, η_i is the unobserved individual effect with zero mean $E(\eta_i) = 0$, and the transitory component ν_{it} is assumed to have finite moments and is serially uncorrelated (i.e. $E(\nu_{it}\nu_{is}) = 0$ for $t \neq s$). We consider small T and large N asymptotics. We assume that $E(\nu_{it}|d_i, \eta_i, y_{i0}) = 0 \quad \forall t$, which implies that several standard assumptions concerning error components models hold. In particular, it implies $E(\nu_{it}) = 0$ and $E(\eta_i\nu_{it}) = 0$ for all t, and the standard assumption concerning the initial condition holds, i.e. $E(y_{i0}\nu_{it}) = 0$ for all $t = 1, ..., T$ (see for example Ahn and Schmidt (1995)). It also implies that d_i is strictly exogenous (i.e. $E(d_i\nu_{it}) = 0 \quad \forall t$). However, no restrictions are placed on the correlation between d_i and η_i. It is straightforward to show that the following moment restrictions hold:

$$E(y_{i,t-1}\nu_{is}) = 0, \qquad E(d_i y_{i,t-1}\nu_{is}) = 0 \qquad \forall s \geq t, t = 1, ..., T \tag{4}$$

The key difference of this model from existing dynamic models of panel data is that the autoregressiveness depends on the dummy variable d_i. d_i will help us identify α_{1d}, the key parameter of interest.

We consider identification in a linear "difference GMM" setting without using the strictly exogenous regressors X_{it}. Both the Anderson and Hsiao (1981a) (AH) and Arellano and Bond (1991) (AB) estimators require taking first differences, then use two lags or more of the dependent variable as instruments in the

[19]In a separate paper, the author develops a related model where d is allowed to vary over time. In this paper, it suffices to treat d as constant over time.

first-difference equations. We follow their approach by taking first-differences:

$$y_{it} - y_{i,t-1} = \alpha_1(y_{i,t-1} - y_{i,t-2}) + \alpha_{1d}(d_i y_{i,t-1} - d_i y_{i,t-2}) + \Delta\nu_{it} \qquad t = 2, ..., T \tag{5}$$

where $\Delta\nu_{it} \equiv \nu_{it} - \nu_{i,t-1}$. Condition (4) allows the use of $y_{i,t-2}$ and further lags as instruments in the first-difference equations. In the AH and AB style estimators, the moment restrictions that can identify α_1 are $E(y_{i,t-2-s}\Delta\nu_{it}) = 0 \quad \forall s = 0, ..., t-2$. Similarly, our moment restrictions that can identify α_{1d} are:

$$E(d_i y_{i,t-2-s}\Delta\nu_{it}) = 0 \qquad \forall s = 0, ..., t-2 \tag{6}$$

4.2 GMM Estimator

We proceed with setting up the instrument matrix in the style of Anderson and Hsiao (1981a). The instrument matrix is the key distinction between our estimator and others. For notational simplicity, let us consider the case where only the second and the third order lagged dependent variables (i.e. $y_{i,t-2}, y_{i,t-3}$) are used as instruments in the first-difference equations. The sample moments of the AH-type estimator for individual i are $\sum_{t=2}^{T} y_{i,t-2}\Delta\nu_{it}$ and $\sum_{t=3}^{T} y_{i,t-3}\Delta\nu_{it}$.[20] We add the sample moments $\sum_{t=2}^{T} d_i y_{i,t-2}\Delta\nu_{it}$ and $\sum_{t=3}^{T} d_i y_{i,t-3}\Delta\nu_{it}$. The instrument matrix for individual i is $Z_i = \left[\begin{array}{cc|c} Z_{iy} & d_i Z_{iy} & 0 \\ \hline 0 & 0 & Z_{ix} \end{array}\right]$, where

$$Z_{iy} = \begin{bmatrix} y_{i0} & 0 \\ y_{i1} & y_{i0} \\ \vdots & \vdots \\ y_{i,T-2} & y_{i,T-3} \end{bmatrix}_{(T-1)*2} \qquad Z_{ix} = \begin{bmatrix} X_{i1} \\ \vdots \\ X_{iT} \end{bmatrix}_{T*K_x} \tag{7}$$

$d_i Z_{iy}$ is the product of the scalar d_i and the matrix Z_{iy}. The first and second columns of Z_{iy} use two and three lags of y as instruments respectively. Together with the instruments $d_i y_{t-2}$ and $d_i y_{t-3}$ that correspond to the first and second columns of $d_i Z_{iy}$, there are 4-2=2 overidentifying restrictions. Z_{ix} contains moments used to identify γ.[21]

The linear GMM estimator is standard (Hansen (1982)) and is sketched below. Let the residual vector be[22] $\tilde{\nu}_i = [\Delta\nu_{i2}...\Delta\nu_{iT}|\nu_{i1}...\nu_{iT}]'$. The residual vector can be expressed in the linear form $\tilde{\nu}_i = \tilde{y}_i - \tilde{X}_i b$,

[20] In contrast, the AB-type estimator treat the terms in the sums as separate moments.

[21] If the within-group moment restrictions on X are used, we have $Z_{ix} = \left[(X_{i1} - \bar{X}_i)'...(X_{iT} - \bar{X}_i)'\right]'$. If the first-difference moment restrictions on X are used, we have $Z_i = \left[\begin{array}{ccc} Z_{iy} & Z_{iy} \circ Z_{id} & Z_{ix} \end{array}\right]$ where $Z_{ix} = \left[\Delta X'_{i2}...\Delta X'_{iT}\right]'$.

[22] If the within-group moment restrictions on X are used, we have $\tilde{\nu}_i = [\Delta\nu_{i2}...\Delta\nu_{iT}|(\nu_{i1} - \bar{\nu}_i)...(\nu_{iT} - \bar{\nu}_i)]'$. If the first-difference moment restrictions on X are used, we have $\tilde{\nu}_i = [\Delta\nu_{i2}...\Delta\nu_{iT}]'$.

where

$$\tilde{y}_i = \left[\begin{array}{c} \Delta y_{i2} \\ \vdots \\ \Delta y_{iT} \\ \hline y_{i1} \\ \vdots \\ y_{iT} \end{array}\right] \tilde{X}_i = \left[\begin{array}{ccc} \Delta y_{i1} & \Delta d_i y_{i1} & \Delta X_{i2} \\ \vdots & \vdots & \vdots \\ \Delta y_{i,T-1} & \Delta d_i y_{i,T-1} & \Delta X_{iT} \\ \hline y_{i0} & d_i y_{i0} & X_{i1} \\ \vdots & \vdots & \vdots \\ y_{i,T-1} & d_i y_{i,T-1} & X_{iT} \end{array}\right] \tag{8}$$

The first order condition of the minimization criteria

$$\min_b \left(\frac{1}{N}\sum_{i=1}^{N} Z_i'\tilde{\nu}_i\right)' A \left(\frac{1}{N}\sum_{i=1}^{N} Z_i'\tilde{\nu}_i\right) \tag{9}$$

with weighting matrix A yields a GMM estimator.

Letting $D_{zx} \equiv \frac{1}{N}\sum_{i=1}^{N} Z_i'\tilde{X}_i$ and $D_{zy} \equiv \frac{1}{N}\sum_{i=1}^{N} Z_i'\tilde{y}_i$, the GMM estimator has a closed form which is

$$b_{GMM} = \left(D_{zx}'AD_{zx}\right)^{-1} D_{zx}'AD_{zy} \tag{10}$$

The one-step estimator uses weighting matrix $A = \left(\frac{1}{N}\sum_{i=1}^{N} Z_i'\tilde{H}Z_i\right)^{-1}$ with $\tilde{H} = \left[\begin{array}{c|c} H & 0 \\ \hline 0 & I_T \end{array}\right]$, where H is a $(T-1)*(T-1)$ matrix with entries $h_{ij} = 2$ if $i = j$, $h_{ij} = 1$ if $|i-j| = 1$, $h_{ij} = 0$ if $|i-j| > 1$.[23] The feasible estimated variance of the one-step estimator is

$$Var(b_{GMM1}) = \frac{1}{N}(D_{zx}'AD_{zx})^{-1}D_{zx}'A\hat{V}AD_{zx}(D_{zx}'AD_{zx})^{-1} \tag{11}$$

where $\hat{V} = \frac{1}{N}\sum_{i=1}^{N} Z_i'\hat{\tilde{\nu}}_i\hat{\tilde{\nu}}_i'Z_i$, $\hat{\tilde{\nu}}_i$ equals the estimated GMM residual vector. The two-step estimator uses the optimal feasible weighting matrix $A = \hat{V}^{-1}$. The feasible estimated variance of the two-step estimator is

$$Var(b_{GMM2}) = \frac{1}{N}(D_{zx}'\hat{V}^{-1}D_{zx})^{-1} \tag{12}$$

The Sargan test statistic of over-identifying restrictions (Sargan (1958)) is given by the value of the objective function (9) evaluated at the 2-step GMM estimator. Asymptotically, it follows a chi-square distribution with degrees of freedom $R-k$ where R is the total number of instruments and k is the number of parameters in the model. Rejection of the null hypothesis implies that some of the instruments are invalid. The m_2 statistic (Arellano and Bond (1991)) tests for serial correlation in the transitory error. Asymptotically, it follows a standard normal distribution. Rejection of the null hypothesis implies that the transitory errors of

[23] In the case where the first difference transformation on X is used, the one-step GMM estimator using $\tilde{H} = H$ will be efficient under conditional homoskedasticity (e.g. Arellano and Bond (1991)).

the dynamic model are serially correlated, and therefore lags of the dependent variable cannot be used as instruments.

Our estimator is different from standard AH and AB estimators because the instrument matrix includes the dummy variable d_i. In addition, as opposed to the original AH estimator, it uses further lags of the dependent variable to form overidentifying restrictions. There are two reasons why the AB style estimator is not used. First, the moment restrictions regarding lags of the dependent variable doubles due to the inclusion of d_i. In finite samples, the use of too many moment conditions causes the Sargan test to be undersized and to have extremely power (Bowsher (2002)). Our benchmark sample contains 20 months of data which easily lead to more than 100 overidentifying restrictions. As indicated in the robustness analysis, the AB style estimator does not perform well in our data. The other reason is for the more general case where d is allowed to vary over time, which tends to be true for survey data. The instrument matrix for the AB style estimator is much harder to set up in that case. For example, the instrument matrix will have less than full rank when $d_{it} = 0$ for all i in some period t.

4.3 Application to Time Limits

The time limit model in equation (1) can be applied to the generic dynamic panel data model in equation (3). Taking first difference in (1), we have

$$\begin{aligned}\Delta y_{it} = \alpha_1 \Delta y_{i,t-1} + \alpha_2 \Delta y_{i,t-2} + \alpha_3 \Delta y_{i,t-3} + \beta_s d_i y_{i,t-1} + \beta_A d_i \Delta A_{it} + \\ \gamma_s y_{i,t-1} + \gamma_A \Delta A_{it} + \Delta X_{it}\gamma + \Delta \nu_{it}\end{aligned} \tag{13}$$

After taking first difference, the terms $\beta_s d_i \Delta S_{it}$ and $\gamma_s \Delta S_{it}$ reduce to $\beta_s d_i y_{i,t-1}$ and $\gamma_s y_{i,t-1}$ respectively. Unlike the example in (2), welfare participation lagged two periods and more can be used as instruments in the first-difference equation. Let R_y be the number of lagged dependent variables used as instruments. There are 3+2=5 endogenous variables on the right hand side, so there are $2 * R_y - 5$ overidentifying restrictions. We use the two-step GMM estimator because it produces more efficient estimates under heteroscedasticity, which is the case in a linear probability model. Instead of using first-difference transformations (e.g. Arellano and Bond (1991)), we use OLS-type moment restrictions for the strictly exogenous regressors because most of them are not truly time-varying.[24]. In addition, a more direct comparison can be made between our estimates and estimates from existing models, which are based on OLS.[25] The instrument matrix is

[24]For example, the age of the youngest child in each period is imputed from the Background Demographic Form, which contains information at the time of random assignment only. The FTP dummy is time-invariant

[25]Mazzolari (2007)'s model is based on instrumental variables. However, the key distinction is that her model does not consider unobserved heterogeneity.

$$Z_i = \left[\begin{array}{cc|cccc} Z_{iy} & d_i Z_{iy} & & 0 & & \\ \hline 0 & 0 & d_i A_i & A_i & d_i & X_i \end{array}\right] \tag{14}$$

where

$$Z_{iy} = \begin{bmatrix} y_{i2} & y_{i1} & y_{i0} & 0 & \dots & 0 \\ y_{i3} & y_{i2} & y_{i1} & y_{i0} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ y_{i,T-2} & y_{i,T-3} & y_{i,T-4} & y_{i,T-5} & \cdots & y_{i,T-1-R_y} \end{bmatrix}_{(T-3)*R_y} \tag{15}$$

4.3.1 Nonlinear Stock Effect

Under an intertemporal optimization model, the incentive to bank welfare increases nonlinearly when a person has received more months of welfare under a time limit. For simplicity, suppose everyone in the program group faces the same type of time limit, so that the stock and months of remaining eligibility are equivalent concepts. The nonlinear incentive to bank welfare can be captured via a adding spline construct in the benchmark model (1):

$$\begin{aligned} y_{it} = \alpha_1 y_{i,t-1} + \alpha_2 y_{i,t-2} + \alpha_3 y_{i,t-3} + \beta_s d_i S_{it} + \beta_{s1} d_i \max\{S_{it} - s_1, 0\} + \beta_A d_i A_{it} + \\ \gamma_s S_{it} + \gamma_{s1} \max\{S_{it} - s_1, 0\} + \gamma_A A_{it} + \gamma_d d_i + X_{it}\gamma + \eta_i + \nu_{it} \qquad t = 3, ..., T \end{aligned} \tag{16}$$

Once the total time on welfare under a time limit reaches the threshold s_1, for each additional month on welfare, the family's probability of welfare receipt drops by $\beta_s + \beta_{s1}$. The spline specification can be used to test the stock effect hypothesis in greater detail. We use splines instead of the apparently simpler quadratic form (S_{it}, S_{it}^2) because the quadratic form is not identified due to the endogeneity of S_{it}.

The spline is identified by two extra moment conditions. The instrument matrix for the GMM estimator for this model is:

$$Z_i = \left[\begin{array}{cccc|cccc} Z_{iy} & d_i Z_{iy} & \sum_{j=1}^{R_y} Z_{iys}[*j] & d_i \sum_{j=1}^{R_y} Z_{iys}[*j] & & 0 & & \\ \hline 0 & 0 & 0 & 0 & d_i A_i & A_i & d_i & X_i \end{array}\right] \tag{17}$$

where Z_{iys} is the entrywise product of Z_{iy} and Z_{is} (i.e. $Z_{iys} \equiv Z_{iy} \circ Z_{is}$), Z_{is} is a matrix with indicator functions as entries

$$Z_{is} = \begin{bmatrix} \mathbf{1}(S_{i2} > s_1) & \mathbf{1}(S_{i1} > s_1) & \mathbf{1}(S_{i0} > s_1) & 0 & \dots & 0 \\ \mathbf{1}(S_{i3} > s_1) & \mathbf{1}(S_{i2} > s_1) & \mathbf{1}(S_{i1} > s_1) & \mathbf{1}(S_{i0} > s_1) & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \mathbf{1}(S_{i,T-2} > s_1) & \mathbf{1}(S_{iT-3} > s_1) & \mathbf{1}(S_{i,T-4} > s_1) & \mathbf{1}(S_{i,T-5} > s_1) & \dots & \mathbf{1}(S_{i,T-1-R_y} > s_1) \end{bmatrix}_{(T-3)*R_y} \tag{18}$$

and $\sum_{j=1}^{R_y} Z_{iys}[*j]$ is the sum of all columns in Z_{iys}, producing a vector. This vector is nonzero whenever S is larger than the threshold s_1 at some time t. The reason for summing the columns is that columns to the right hand side of Z_{is} could contain many or even all zero entries. The moment conditions formed by these columns contain less information than the moment conditions formed by columns to the left hand side of Z_{is}. The number of overidentifying restrictions remains unchanged because the two extra moment conditions just identify the two extra parameters.

5 Results

In this section we present estimates for a variety of empirical specifications of time limit models, compare the performances among different models, discuss the goodness-of-fit, and perform simulation exercises.

5.1 Key Estimation Results

Table 2 compares dynamic GMM estimates with OLS and fixed effects estimates. Welfare participation from month 4 to 6, as well as the actual months of welfare receipt since random assignment in month 6, are used as initial conditions in dynamic models.[26] Since the sample period ends in month 23, the results should capture the child and stock effects but not any mechanical effects.

Column 1 contains results from the benchmark model in equation (1), estimated by the GMM estimator developed earlier with the instrument matrix defined in (14). The benchmark GMM model uses 12 lagged dependent variables (i.e. $R_y = 12$) and their interaction with the FTP dummy as instruments. The first row reports the stock effect of the time limit. Other things unchanged, for each additional year of welfare receipt under a time limit, the welfare receipt probability drops by 6.4 percentage points. For example, for families which have received 18 months of welfare under a time limit, the time limit caused the welfare receipt probability to drop by 9.6 percentage points. The result is significant at the 1 percent level. The stock effect is ten times as large as the child effect reported in the second row. For each additional year of age of the youngest child, the welfare receipt probability only decreases by 0.6 percent. In terms of the terminology of intertemporal optimization, this suggests that the welfare-banking behavior of families was much more sensitive to the state variable (i.e. total time of welfare receipt) than the time horizon (i.e. age

[26] Section 4 shows that the choice of the initial S does not affect estimation results.

of the youngest child). A numerical example in the appendix shows that a low time discount factor results in a high ratio of the stock effect to the child effect. Both the Sargan and the m2 tests support the null hypothesis that the instruments are valid. Receiving welfare now increases the probability of receipt by 41 percent next month, by 9 percent the month after, and decreases the probability by 3 percent in the third month. [27]

Column 2 considers a model with unobserved heterogeneity but no state dependence (i.e. no lagged welfare participation). The stock is still considered endogenous. The Sargan test rejects the null hypothesis at the 1 percent level, indicating that the model is misspecified. Omitting state dependence creates a large upward bias (about 50 percent) in the magnitude of the stock effect. The model in column 3 has unobserved heterogeneity but omits both the state dependence and the stock effect. Omitting the stock effect creates a downward bias in the child effect. The reason is explained as follows. Families with younger children accumulate the stock slower due to the child effect. But as time passes, the lower stock positively affects welfare receipt due to the stock effect. Therefore the estimated child effect becomes smaller when the stock variable is not controlled for. This is consistent with Mazzolari (2007)'s observation, which states that if the stock is miscalculated, the model becomes more misspecified as time passes.[28]

Columns 4 and 5 report results from OLS. Both OLS models do not have unobserved heterogeneity. Column 4 has state dependence while column 5 does not have state dependence. Omitting unobserved heterogeneity creates a large downward bias (about 50 percent) in the magnitude of the stock effect. This result is not surprising. Families with high unobserved effects have a higher stock and a higher welfare receipt probability. This will offset the true stock effect if unobserved heterogeneity is not controlled for in the model.

Columns 6 and 7 report results from the simple fixed effects (FE) model, which accounts for unobserved heterogeneity. Column 6 has state dependence while column 7 does not have state dependence. The simple FE model is misspecified because it assumes all right hand side variables are strictly exogenous (i.e. they are uncorrelated with past, concurrent and future transitory errors). In particular, since lagged welfare participation and S_{it} are positively correlated with past transitory errors,[29] the corresponding coefficients are upward biased. Indeed, the estimated state dependence is higher, and the magnitude of the stock effect becomes lower.[30]

[27] For the control group, the welfare receipt probability drops by 23 percent for each additional year on welfare since random assignment. This could be partly due to the greater welfare reform environment, or because control group members erroneously believed that they were subject to a time limit. We do not construct a behavioral model to explain this result, but we include this term in the empirical model to adequately control for the behavior of the control group.

[28] [...the characterization of the effects of time limits worsens as time passes since the time limit is introduced].

[29] Correlation is allowed between the lagged dependent variables and the unobserved effect, so the FE model is less misspecified than the the RE model.

[30] The simple FE model uses within-group transformations to estimate the parameters. Results regarding the FTP dummy and the age of the youngest child are not reported because of inadequate within-group variations.

5.2 Within-Sample Fit and Extrapolation

We first compare the our GMM model with the Grogger and Michalopoulos (2003) and Mazzolari (2007) models[31] in terms of within-sample fit and extrapolation. We present the predicted and actual outcomes for the program group from month 7 to 53. The key outcome is the welfare participation rate, and we also discuss the benefit exhaustion rate, defined as the percentage of the program group that has reached the time limit cap (which is 24 or 36 months). As discussed in section 2, the sample period of the models ends in month 23 to avoid dealing with mechanical effects in the estimation process. Therefore, predictions starting from month 24 are extrapolations and can be used to assess the predictive performances of the models.

The GM model considers only the child effect, and the Mazzolari model imputes the stock using pre-time-limit welfare participation rates. Details of their models are given in the appendix. Since both models do not control for unobserved effects and are static (i.e. no lagged welfare participation), the predicted outcomes are relatively straightforward to compute. The benchmark GMM model that we use comes from column 1 in Table 2.[32] In the GMM model, we first use the residuals to estimate the unobserved effect of each family, then use the unobserved effects, estimated coefficients, and actual welfare receipt from month 4 to 6 as initial conditions to predict the welfare participation rate. In all three models, the predicted stock is also computed. When a family's predicted stock reaches the time limit cap, predicted welfare receipt is set to zero thereafter. In the data, due to exemptions and extensions, a small fraction of families continue to receive benefits after reaching the time limit cap. In order to compare the actual and predicted results more consistently, for families that have reached the cap, we redefine actual welfare receipt in all subsequent periods to zero.

The GMM, GM and Mazzolari models are compared in Figure 2. The GM and Mazzolari models fail to capture the continuing drop of the welfare participation rate after random assignment. The major cause is that there is no state dependence (i.e. lagged welfare participation) in both models. As a result, both models over-predict the welfare participation rate starting from month 24. Even so, both models predict that the first family does not exhaust its benefits until month 52, which is almost 30 months after the first family actually exhausts its benefits. The major cause of the under-prediction of the benefit exhaustion rate is that both models do not incorporate unobserved heterogeneity. Under unobserved heterogeneity, families with high unobserved effects will exhaust their benefits more quickly. In contrast, the predicted welfare participation rate of the GMM model matches the actual rate closely both within and outside the sample period. Figure 3(a) reports the actual and predicted benefit exhaustion rates as percentages of the program group. The GMM model successfully predicts that some families which are subject to the 24-month limit exhaust their benefits at month 24. However, it does not predict well that some families which are subject to

[31] The Fang and Keane (2004) model is not identified by FTP data.

[32] The GMM, GM and Mazzolari models used for prediction are based on models that exclude time dummies and keep the age of the youngest child time-invariant. Time effects are kept minimal.

the 36-month limit exhaust their benefits in month 36. Figure 3(b) compares the actual and the predicted stock of the program group at the 25th percentile, mean and 75th percentile. The GMM model matches not only the mean of the stock but also its distribution both within and outside the sample period.[33]

5.3 Counterfactual Exercises

5.3.1 Treatment Effect and its Decomposition

We perform a counterfactual exercise on the program group to examine the "treatment effect of the treated". Here the treatment refers to the time limit alone, not FTP as a whole. The simulation method has been discussed in the previous subsection. First, we simulate the welfare participation rate from month 7 to 53 assuming the time limit is in effect. Then, we simulate welfare participation assuming there is no mechanical effect, i.e. families can continue to receive benefits after reaching the cap. We then simulate welfare participation assuming no mechanical and stock effects (by setting $\gamma_s = 0$). Finally, we simulate welfare participation assuming no mechanical, stock and child effects (by setting $\gamma_A = 0$). We consider the last case as "no treatment". The difference between treatment and no treatment is the treatment effect. The second and the third cases decompose the treatment effect and allow us to analyze the relative contributions of the child, stock and mechanical effects.

Figure 4 reports the results of the treatment effect and its decomposition. From month 7 to 23, the time limit causes welfare participation to drop by an average of 29 percent for the program group. This is considerably larger than Grogger and Michalopoulos (2003)'s estimate of 16 percent, which considers only the child effect using similar data from FTP. The child and the stock effects constitute 60 and 40 percent of the drop respectively. This suggests that the difference in the estimates is related to the omission of the stock effect in GM's model. The contribution of the stock effect rises from almost 0 percent in month 7 to 50 percent in month 23. From month 24 to 53, the time limit causes welfare participation to drop by an average of 45 percent. The child, stock and mechanical effects constitute 35, 50 and 15 percent respectively. The mechanical effect is much smaller than Mazzolari (2007)'s result from the Survey of Income and Program Participation (SIPP), which finds that the mechanical effect constitutes 80 percent of the total effect of the time limit on welfare receipt from 1996 to 2003. When we shut off the stock effect in the counterfactual exercise, the share of the mechanical effect jumps to 60 percent. Therefore, Mazzolari's model may have attributed too much of the decline in welfare receipt to the mechanical effect because of stock imputation.

5.3.2 Simulating Lifetime and Periodic Time Limits

We assume that welfare participation behavior is at a steady state and simulate the effects of two types of time limits: 1) a lifetime (permanent) 24-month limit; 2) a limit of a maximum of 24 months in any

[33] We also tried setting unobserved effect as zero for each family when performing simulation. Results on welfare participation do not differ greatly, but the predicted 75th percentile of the stock becomes much smaller than the actual counterpart.

60-month period. The steady-state assumption is interesting because we can control for systematic changes in the external environment. The steady state participation rate is set at 50 percent, which is the average rate during the 23-month period before the random assignment. The first and second order autoregressive parameters are set at 0.4 and 0.1 respectively, which are estimates from the benchmark model. We assume that the total time of welfare receipt does not have an effect on welfare participation (i.e. $\gamma_s = 0$) so that the steady state is defined. The stock effect (β_s) is set to be 6 percent per year. The child effect (β_A) is set to be 0.6 percent per year. The unobserved effect of each family is estimated from the benchmark GMM model. We assume that the age of the youngest child does not vary over time.

Figure 5(a) reports the simulation results for the 24-month lifetime limit. With the child effect only, welfare participation drops by 35 percent within the first year and stays constant thereafter. With the child, stock and mechanical effects, welfare participation drops continually. Within the first 5 years, the welfare participation rate drops by 70 percent. Note that the simulation is based on estimates from families which are subject to non-lifetime limits. Therefore, the actual effect of the lifetime limit may be larger. Figure 5(b) reports the simulation results for the 24-of-60 month periodic limit. This policy restricts welfare receipt to a maximum of 24 months in any 60-month period. The child effect exhibits the same magnitude. With the child, stock and mechanical effects, the welfare participation rate drops to 15 percent (i.e. a reduction of 70 percent) by the end of the 5th year, jumps back to 20 percent during the 6th year, and stays roughly constant thereafter. There are two important differences. First, welfare participation reaches to a new steady state at around 20 percent, while in the case of a lifetime limit, a new steady state does not exist. Second, the mechanical effect is much smaller. The main reason is that the stock effect acts as a buffer and prevents people from reaching the time limit cap.

5.4 Program Subgroups and Nonlinear Stock Effects

In FTP, approximately half of the program group members are allowed a maximum of 24 months of welfare receipt within any 60-month period. The other half are considered as the "particularly disadvantaged" and are allowed a maximum of 36 months of welfare receipt within any 72-month period. In this subsection, we find that the two subgroups react similarly to the time limit(i.e. the treatment effect does not seem to be heterogeneous). Using spline specifications, we find that the incentive to bank benefits is only significant when families approach the time limit cap.

The 36-month limit group, being more disadvantaged, has a higher welfare participation rate than the 24-month limit group. From month 0 to 23, the welfare participation rate of the 36-month limit group is on average 30 percentage points higher than the 24-month limit group. In month 24, 68 percent of the 36-limit group have received welfare for more than 12 months since random assignment. Only 6 percent of the group never received any welfare. In contrast, in month 24, only 37 percent of the 24-limit group have received welfare for more than 12 months since random assignment. Around 18 percent of the group never received any welfare. It is hardly convincing that the 36-month limit group receives more welfare just because the observable characteristics are different. It is also likely that the 36-month limit group has higher unobserved effects on average.

There are two complications involved when we decompose the program group. First, the control group has to be treated as a whole because it does not make a distinction between the particularly disadvantaged and the non-disadvantaged groups. Hence, we construct two samples: A) control group and the 24-month limit group covering month 4 to 23 from random assignment; B) control group and the 36-month limit group covering month 4 to 35 from random assignment. We control for observable characteristics as well as unobserved effects. We allow state dependence to differ between the control group and the program subgroups. Second, the two program subgroups may react differently just because they are subject to different time limits. However, for both the 24-of-60 and 36-of-72 month limits, families are ineligible for 36 months in a row if they have received welfare for 24 or 36 months in a row. This suggests that the effects should be similar for families that have the same number of months of remaining eligibility.[34]

Table 3 reports estimation results of the GMM model for samples A and B above. Columns 1 to 2 report results from sample A. Column 1 shows that the average stock effect for the 24-month limit subgroup is 30 percentage points per additional year of welfare receipt. Although the magnitude seems very large, it should be noted that the 24-month limit group receives much less welfare than the 36-month limit group on average. Columns 2 reports results for the model with one spline at the 12th month of welfare receipt (i.e. 1 year of eligibility remaining). The stock effect from the time limit is only significant at the spline. For families with less than 1 year of eligibility remaining, the stock effect is 17 percentage points per additional year of welfare receipt. For the control group, there is no similar non-linearity pattern. Columns 3 to 4 report results from sample B. Column 3 shows that average stock effect for the 36-month limit subgroup is 8.4 percentage points per additional year of welfare receipt. Columns 4 reports results for the model with one spline for the stock effect at the 24th month of welfare receipt (i.e. 1 year of eligibility remaining). Again, the stock effect from the time limit is only significant at the spline. For those with less than 1 year of eligibility remaining, the stock effect is 20 percentage points per additional year of welfare receipt. For the control group, there is no similar non-linearity pattern.

[34] Consider two hypothetical cases. In the first case, suppose in month 23, an individual subject to the 24-month limit has 1 month of eligibility remaining. If she receives welfare in month 24, she will be ineligible for 36 months until month 60. In the second case, suppose in month 35, an individual subject to the 36-month limit has 1 month of eligibility remaining. If she receives welfare in month 36, she will be ineligible for 36 months until month 72. Therefore, the behavior of both families should be roughly the same in both cases, given all other characteristics identical.

5.5 Robustness Analysis

We perform a number of robustness checks on the estimation results. The estimates from the benchmark model are not sensitive to the exclusion of time dummies, as well as allowing the covariates to be correlated with unobserved effects. Table 4 reports results from two types of robustness checks. The first type is to examine if the estimates are sensitive to the sample period. Column 1 contains results from the benchmark model estimated using the 4th to the 23rd month BEFORE random assignment. There are no stock and child effects before random assignment, indicating that the stock and child effects detected after random assignment should be due to policy differences but not to systematic differences between the program and control groups. Column 2 contains results from the benchmark model estimated using month 0 to 23. The first 4 months are no longer discarded. The stock and the child effects are about half as large but are still significant. However, the Sargan test rejects the null hypothesis that the instruments are valid. This suggests that a different dynamic process (such as program implementation constraints) might be in effect in the first few months after random assignment. Column 3 reports results from the benchmark model estimated using an unbalanced panel constructed from a longer sample period. For control group members, month 4 to 53 are included. For program group members, the sample period starts from month 4 and ends in month 53 or the month before the benefit is exhausted, whichever is earlier. The stock effect remains roughly the same, while the child effect becomes insignificant. However, both the Sargan and the m2 test reject the null hypothesis that the model is correctly specified.

Columns 4 to 6 report results obtained using different econometric specifications. The benchmark model uses 12 lagged dependent variables and their interactions with the FTP dummy to form the instrument set. Column 4 and column 5 use 18 and 5 lagged dependent variables respectively. The stock and child effects remain similar, suggesting that our results are not sensitive to the size of the instrument set. Column 6 uses an Arellano and Bond (1991) type estimator with 5 lagged dependent variables.[35] There are 149 overidentifying restrictions. The stock effect becomes insignificant, but the Sargan test rejects the null hypothesis that the instruments are valid.

6 Conclusions

Time limits enjoy policy appeal because it is argued that welfare recipients quit proactively before they reach the time limit cap. Although time limits have already been widely implemented in the United States, they are still under policy scrutiny in a number of other countries. An important mechanism of the time limit is the stock effect, defined as the increased incentive to bank welfare benefits for future use when the total months of welfare receipt under the time limit increases. The welfare literature has yet developed

[35] This estimator uses within-group transformation on the strictly exogenous variables. Coefficients involving the FTP dummy and the age of the youngest child are not reported because they do not exhibit adequate variation over time.

a model that can satisfactorily estimate the stock effect. Our paper fills this gap by developing a linear dynamic panel data model and use data from Florida's FTP policy experiment to estimate the stock effect. FTP has a relatively large sample size (2570 in our sample) and a long time period (from 2 years before random assignment to 4.5 years after random assignment), allowing the dynamic model to be estimated with reasonable precision. Our estimates suggest that the stock effect is sizable, and omitting it will lead to a severe misrepresentation of the true impacts of time limits. During the first two years from random assignment, the FTP time limit caused welfare participation to drop by an average of 29 percent. Around 40 percent of the drop was due to the stock effect.

These results have implications for the welfare reform experience in the United States. First, recent work on time limits which model the child effect such as Grogger (2003), Grogger (2004) and Fang and Keane (2004) find that time limits explained around 11 to 13 percent of the decline in welfare from the late 1990s to the early 2000s. If our result regarding the stock effect is generalizable to the nation as a whole, the proactive welfare-quitting behavior induced by time limits should have explained around 20 to 25 percent of the decline. This implies that the effects of welfare reform should have been stronger than found in previous studies using either aggregate-level data (e.g. U.S. Council of Economic Advisors (1997), U.S. Council of Economic Advisors (1999), Moffitt (1999), Schoeni and Blank (2000), Ziliak, Figlio, Davis, and Connolly (2000)) or individual-level data (e.g. Fang and Keane (2004), Grogger (2004)).[36] Second, existing estimates of the mechanical effect could be too large. For example, Mazzolari (2007) finds that 80 percent of the decline in the welfare participation rate in the US from 1996 to 2003 were due to the mechanical effect. Our work suggests that, due to the stock effect, the welfare participation rate should have become very low for families approaching the time limit cap. Therefore, the incidence of reaching the cap for these families by itself should not cause a large drop in the welfare participation rate of the larger population.

Two lessons can be learned from this analysis. First, existing national estimates of the behavioral effects of time limits in the US could be biased downwards because the stock effect is not fully accounted for. As a result, we expect that a larger fraction of the unprecedented drop in the welfare participation rate following welfare reform can be attributed to time limits. Second, the mechanical effect may have been overstated in existing studies. Due to the stock effect, welfare participation drops to a very low level when families approach the time limit. Putting aside the potential benefits of proactively leaving welfare (such as building human capital through work), this suggests that the well-being of families reduces substantially well before they reach the time limit cap.

[36]Mazzolari (2007) attributes 25 percent of the decline in welfare use from 1996 to 1999 to time limits. However, she finds that 80 percent was due to the mechanical effect.

A Appendix

A.1 Some Existing Models of Time Limits

This section outlines the Grogger and Michalopoulos (2003) and Mazzolari (2007) models, which are closely related to our model. We make simplifications to both models but keep the essential components. Predicted outcomes are computed from the adapted models. The GM model is:

$$y_{it} = \gamma_d d_i - \beta_A d_i \max\{18 - A_{it} - \bar{S}_i, 0\} + X_{it}\gamma + \epsilon_{it} \tag{19}$$

where d_i is the FTP dummy, A is the age of the youngest child in the family (in years), $\bar{S}_i$ is the time limit length (in years), X is a set of exogenous variables, ϵ is the error term. β_A captures the child effect. $\max\{18 - A_{it} - \bar{S}_i, 0\}$ equals zero whenever the time limit is not binding.[37] As in GM, we set $\bar{S}_i$ equal to 2 (i.e. 24 months). Estimation based on our benchmark dataset finds that β_A equals 0.9 percent, which is slightly higher than GM's estimate of 0.7 percent.

We consider a simpler version of the Mazzolari model by omitting the part that estimates the mechanical effect. The model is:

$$\begin{aligned} y_{it} &= \gamma_d \mathbf{1}\{18 - A_{it} > \tilde{S}^j_{it}\} d_i + \beta_{AS} \mathbf{1}\{18 - A_{it} > \tilde{S}^j_{it}\} d_i \frac{\tilde{S}^j_{it}}{18 - A_{it}} + X_{it}\gamma + \epsilon_{it} \\ \tilde{S}^j_{it} &= \bar{S}_i - k_j(t - t_i^{TL}) \end{aligned} \tag{20}$$

If $18 - A_{it} > \tilde{S}^j_{it} > 0$ then the time limit is binding. $\tilde{S}^j_{it}$ is the imputed months of remaining eligibility under a time limit and is defined only for the program group. $\bar{S}_i$ is the time limit length, set at 2 or 3 (i.e. 24 or 36 months). $t - t_i^{TL}$ is the time passed since the time limit was put into effect. k_j is the average welfare participation rate from month 1 to 23 before random assignment. It differs between the control and the program group (i.e. $j = 1, 2$). β_{AS} captures a combination of the stock and child effects.

A.2 A Formal Model of Time Limits - the Role of the Time Discount Factor

We use a generic numerical example to show that the time discount factor determines the relative size of the child and stock effects of a time limit. When the discount factor is low (i.e. individuals are less forward-looking), the stock effect becomes much larger than the child effect. Therefore, empirical findings on the relative size of the stock to the child effect may shed some light on how forward-looking individuals are.

Suppose in each period a single woman with a child receives a high wage $\bar{w} > 0$ with probability θ, or a low wage $\underline{w} \geq 0$ with probability $1 - \theta$. At the low wage, she is eligible for a welfare payment $\bar{c}$. At the high wage, she is eligible for a welfare payment $\underline{c}$, where $\underline{c} < \bar{c}$. Let the utility function be linear as

$$u(w, p; \underline{c}, \bar{c}) = \begin{cases} \underline{w} + \bar{c}p & \text{if } w = \underline{w} \\ \bar{w} + \underline{c}p & \text{if } w = \bar{w} \end{cases} \tag{21}$$

where $p \in \{0, 1\}$ denotes welfare participation choice. For simplicity, assume that she works for one time unit and consumes all income each period. There are no marriage and fertility decisions, and there is no welfare stigma.

If the model is static, the person will always participate. However, the person will not always participate if a time limit is introduced. Suppose the individual faces a lifetime time limit of length $\bar{S}$ (i.e. she is allowed a maximum of

[37] GM argues that the child effect is not identified for families with $A < 3$ due to further policy differences between the program and control groups. We exclude these families in the calculation of the child effect by setting $\max\{18 - A_{it} - \bar{S}_i, 0\} = 0$ for these families.

$\bar{S}$ months of cash assistance in her lifetime). Let S_t be the total months of welfare receipt under a time limit at time t. If $S_t = \bar{S}$, the individual is ineligible for welfare forever. Let T be the time until the youngest child in the family turns 18, after which the individual is no longer eligible for welfare. The individual maximizes her expected stream of utility with a discount factor ρ:

$$\max_{\{p_t|w_t\}_{t=0}^{H}} \sum_{t=0}^{T} \rho^t E_0 u(w_t, p_t; \underline{c}, \bar{c}) \qquad \text{s.t.} \qquad S_{t+1} = S_t + p_t \tag{22}$$

This choice problem is similar to Fang and Keane (2004), and is simpler than Grogger and Michalopoulos (1999), who consider joint welfare and work choice. Unlike both models, this model specifies a utility function and a wage distribution for numerical simulation. Risk-neutrality of the individual makes the model more generic, as the numerical results become largely invariant to the magnitudes of $\bar{w}, \underline{w}, \bar{c}$ and $\underline{c}$, or the introduction of more wage outcomes. [38]

At time t, the option value of preserving one month of welfare eligibility is defined as:

$$\mathcal{O}(S_t, t; \rho) \equiv \rho \left(E_t V_{t+1}(S_t) - E_t V_{t+1}(S_t + 1) \right) \tag{23}$$

where V is the value function computed by backwards recursion. The higher is the option value, the lower is the probability of welfare receipt. The option value is a nonlinear function of S and t. Fang and Keane (2004) and Grogger and Michalopoulos (1999) show that (1) the option value is a decreasing function of t, capturing the child effect ($\frac{\partial \mathcal{O}}{\partial t} < 0$, i.e. larger effect on families with younger children) ; (2) the option value is an increasing function of S, capturing the stock effect ($\frac{\partial \mathcal{O}}{\partial S} > 0$, i.e. larger effect on families with higher total time on welfare receipt). The discount factor affects the option value both directly through the multiplicative term and indirectly through the value function.

We linearize the option value function and compare its derivatives under different values of ρ and θ. First, given fixed ρ and θ, we simulate the option value for each combination of S and t. We first choose $\bar{S} = 24$, $T = 216$, $\bar{w} = 10, \underline{w} = 5, \bar{c} = 3, \underline{c} = 1.5$, resulting in $(24-1)*(60-1) = 1357$ data points. Then, we regress the option values on S and t with OLS, resulting in a linearized function:[39]

$$\hat{\mathcal{O}} = c - \hat{\beta_A} t + \hat{\beta_S} S \tag{24}$$

We compute $\frac{\hat{\beta_A}}{\hat{\beta_S}}$, the ratio of the child to the stock effect. We repeat the same analysis with $\bar{S} = 36$.

Figure 6 plots the relationship between this ratio and the time discount factor, under three different probabilities of obtaining a high wage. When the discount factor is low (i.e. individuals are less forward-looking), the stock effect is larger than the child effect, suggesting that behavior is more sensitive to the total time of welfare receipt than the age of the youngest child. This result may not be surprising. When the discount factor is low, the value function converges more quickly in the Bellman equation. When the value function converges, its value depends on the state variable S but not the time unit t. When the discount factor close to 1, the value function converges slowly in the Bellman equation. Therefore, the value function still depends on t even when t is much smaller than the time horizon T.

[38] In addition, under risk-neutrality, the necessary conditions for a time limit to have an effect on behavior are that there are at least two wages where the individual is eligible for welfare, and the welfare benefit decreases strictly with wage. Otherwise, the individual will never bank welfare for future use. When individuals are risk-averse, there should be at least two wages where the individual is eligible for welfare, but the welfare benefit can stay constant with wage.

[39] Discarding data points where the time limit is non-binding only affect the results more when both $\bar{S}$ and T are very small.

Table 1: Summary Statistics at the Time of Random Assignment

	Control	Program			Total
		All	24-month limit	36-month limit	
Mother black (percent)	52.5	52.9	41.9	66.4	52.7
	(49.9)	(49.9)	(49.3)	(47.2)	(49.9)
Mother's age	29.7	29.4	30.3	28.3	29.5
	(7.4)	(7.5)	(7.5)	(7.4)	(7.5)
Years of schooling	11.1	11	11.2	10.7	11.0
	(1.4)	(1.6)	(1.5)	(1.6)	(1.55)
Number of children under 18	2.0	1.9	1.7	2.2	1.9
	(0.9)	(1.0)	(0.9)	(1.0)	(1.0)
Age of youngest child	5.0	4.9	5.3	4.5	5.0
	(4.3)	(4.2)	(4.3)	(4.0)	(4.3)
Total months of welfare receipt	12.2	12.0	8.3	16.5	12.1
in 23 months before RA	(9.1)	(9.3)	(8.4)	(8.3)	(9.2)
Total months of welfare receipt	12.3	12.6	9.6	16.2	12.5
in 23 months after RA	(8.5)	(8.3)	(7.6)	(7.6)	(8.4)
Number of persons	1,286	1,284	708	576	2,570

"RA" denotes the time of random assignment. Table entries are sample means. Standard errors are in parentheses.

Table 2: GMM, OLS and Fixed Effects Estimates of the Effects of Time Limits on Monthly Welfare Participation

	GMM1	GMM2	GMM3	OLS1	OLS2	FE1	FE2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
FTP dummy*Years on Welfare	-0.064***	-0.106***		-0.020***	-0.052***	-0.025**	-0.021
since RA (Annualized)	(0.022)	(0.037)		(0.000)	(0.012)	(0.012)	(0.016)
FTP dummy*Age of Youngest	0.006*	0.008	0.006	0.001*	0.004*	-	-
Child (Annualized)	(0.004)	(0.007)	(0.007)	(0.000)	(0.003)		
Lagged participation (1 month)	0.409***			0.688***		0.546***	
	(0.050)			(0.009)		(0.005)	
Lagged participation (2 months)	0.085***			0.157***		0.115***	
	(0.015)			(0.010)		(0.006)	
Lagged participation (3 months)	-0.027***			0.009		-0.032***	
	(0.008)			(0.007)		(0.005)	
Years on Welfare since RA	-0.228***	-0.697***	-0.751***	0.045***	0.601***	-0.199***	-0.119***
(Annualized)	(0.053)	(0.031)	(0.025)	(0.000)	(0.008)	(0.009)	(0.011)
FTP dummy	0.032	0.011	-0.060	0.011	0.025	-	-
	(0.041)	(0.084)	(0.08)	(0.007)	(0.025)		
Age of Youngest Child (Annualized)	-0.008***	-0.013**	-0.012**	-0.002***	-0.009***		
	(0.004)	(0.006)	(0.006)	(0.000)	(0.002)	-	-
Sargan test	17(19)	115(22)***	122(22)***				
m2 statistic	-0.33	0.09	0.09				
R-squared				0.76	0.40	0.47	0.00
Sample period (RA as month 0)	4-23	4-23	4-23	4-23	7-23	4-23	7-23
Persons	2,570	2,570	2,570	2,570	2,570	2,570	2,570
Observations	43,690	43,690	43,690	43,690	43,690	43,690	43,690

1) "RA" denotes the time of random assignment. Annualized numbers are defined as original, monthly level coefficient estimates (standard errors) times 12. In results (1), (2), (3), (4) and (6), the first three months in the sample period are not directly used due to lagged dependent variables.
2) Results (1) and (2) report two-step GMM estimates using the AH-type estimator developed in the estimation section. 12 lagged dependent variables are used as instruments. The degrees of freedom in the Sargan test of overidentifying restrictions are in parentheses. The m2 statistic refers to the Arellano-Bond (1991) specification test of second-order serial correlation in errors.
3) Standard errors are in parentheses. *, **, *** denote significance at the 10, 5, 1 percent levels respectively.
4) For GMM and OLS results, variables not reported include race, age, years of schooling, number of children, youngest child under 3 dummy, FTP dummy*youngest child under 3 dummy, 36-month time limit dummy, year dummies, and a constant. For fixed effects (FE) results, variables not reported include year dummies and a constant.

Table 3: Decomposing Program Group: GMM Estimates of the Effects of Time Limits on Monthly Welfare Participation

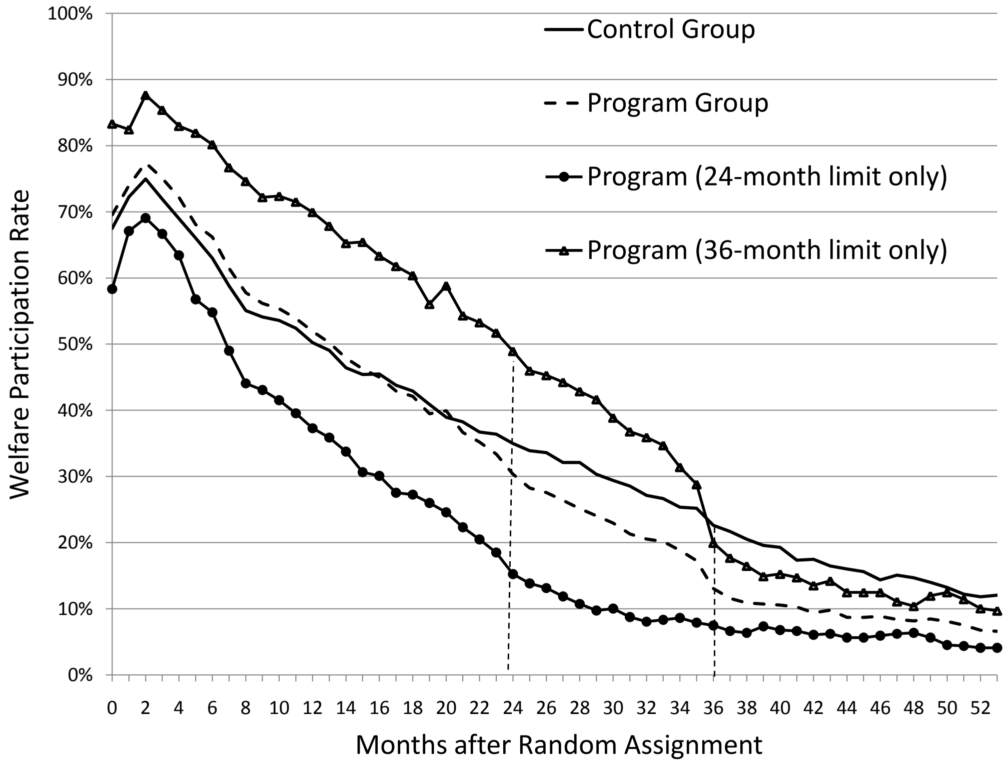
	Control and 24-month Program		Control and 36-month Program	
	(1)	(2)	(3)	(4)
FTP dummy*Years on Welfare since RA (Annualized)	-0.299*	0.146	-0.084*	0.010
	(0.160)	(0.161)	(0.050)	(0.053)
FTP dummy*Years on Welfare since RA Spline 1 (Annualized)		-0.317**		-0.211**
		(0.133)		(0.106)
FTP dummy*Age of Youngest Child (Annualized)	0.007	0.011**	0.002	0.004
	(0.005)	(0.004)	(0.005)	(0.005)
Lagged participation (1 month)	0.423***	0.380***	0.483***	0.479***
	(0.060)	(0.080)	(0.039)	(0.043)
Lagged participation (2 months)	0.101***	0.090***	0.116***	0.115***
	(0.019)	(0.023)	(0.013)	(0.014)
Lagged participation (3 months)	-0.036***	-0.039***	-0.031***	-0.032***
	(0.011)	(0.011)	(0.009)	(0.009)
FTP dummy*Lagged participation (1 month)	-0.036	0.174*	-0.102*	-0.046
	(0.111)	(0.093)	(0.060)	(0.055)
FTP dummy*Lagged participation (2 months)	-0.040	0.000	-0.043*	-0.029
	(0.030)	(0.029)	(0.023)	(0.022)
FTP dummy*Lagged participation (3 months)	0.024	0.030*	0.006	0.010
	(0.017)	(0.018)	(0.016)	(0.016)
Years on Welfare since RA (Annualized)	-0.216***	-0.286**	-0.196***	-0.198***
	(0.061)	(0.113)	(0.042)	(0.055)
Years on Welfare since RA Spline 1 (Annualized)		0.068		0.023
		(0.077)		(0.078)
FTP dummy	0.176	-0.130	0.323***	0.184*
	(0.150)	(0.125)	(0.117)	(0.099)
Age of Youngest Child (Annualized)	-0.008***	-0.010**	-0.007**	-0.007**
	(0.004)	(0.004)	(0.002)	(0.002)
Sargan test	19.6(16)	15.8(16)	20.3(16)	17.9(16)
m2 statistic	-0.05	-0.19	1.16	1.04
Sample period (RA as month 0)	4-23	4-23	4-35	4-35
Persons	1,994	1,994	1,862	1,862
Observations	33,898	33,898	53,998	53,998

1) "RA" denotes the time of random assignment. Annualized numbers are defined as original, monthly level coefficient estimates (standard errors) times 12. The first three months in the sample period are not directly used due to lagged dependent variables.
2) A spline is defined as $\max\{S-c,0\}$ where S is the years on welfare since RA and c is the threshold. In result (2), c equals 1 (equivalent to 12 months). In result (4), c equals 2 (24 months).
3) This table reports two-step GMM estimates using the AH-type estimator developed in the estimation section. 12 lagged dependent variables are used as instruments. The degrees of freedom in the Sargan test of overidentifying restrictions are in parentheses. The m2 statistic refers to the Arellano-Bond (1991) specification test of second-order serial correlation in errors.
4) Asymptotic standard errors are in parentheses. *, **, *** denote significance at the 10, 5, 1 percent levels respectively.
5) Variables not reported include race, age, years of schooling, number of children, youngest child under 3 dummy, FTP dummy*youngest child under 3 dummy, year dummies, and a constant.

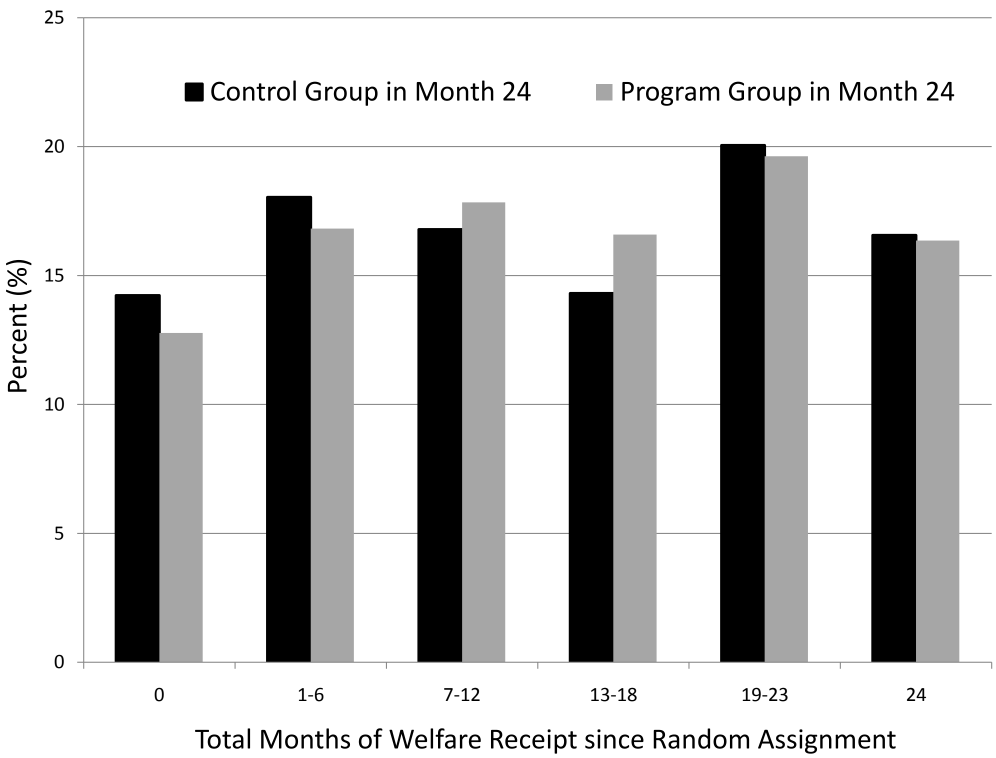
Table 4: Robustness Analysis: GMM Estimates of the Effects of Time Limits on Monthly Welfare Participation

	Change sample period			Change estimation method		
	Pre-RA	Post-RA1	Post-RA2	More instruments	Less instruments	Arellano-Bond estimator
	(1)	(2)	(3)	(4)	(5)	(6)
FTP dummy*Years on Welfare	0.000	-0.035***	-0.059***	-0.065***	-0.059***	-0.037
since RA (Annualized)	(0.001)	(0.012)	(0.014)	(0.022)	(0.023)	(0.023)
FTP dummy*Age of Youngest	0.000	0.002*	0.004	0.007*	0.006*	-
Child (Annualized)	(0.000)	(0.001)	(0.002)	(0.004)	(0.004)	
Lagged participation (1 month)	0.608***	0.628***	0.495***	0.380***	0.430***	0.322***
	(0.053)	(0.037)	(0.022)	(0.049)	(0.085)	(0.039)
Lagged participation (2 months)	0.145***	0.127***	0.107***	0.076***	0.090***	0.064***
	(0.018)	(0.012)	(0.008)	(0.014)	(0.022)	(0.011)
Lagged participation (3 months)	-0.023***	-0.018***	-0.025***	-0.030***	-0.026***	-0.028***
	(0.009)	(0.007)	(0.005)	(0.008)	(0.010)	(0.006)
Years on Welfare since RA	-0.005*	-0.059	-0.186***	-0.265***	-0.208**	-0.406***
(Annualized)	(0.003)	(0.037)	(0.026)	(0.052)	(0.086)	(0.052)
FTP dummy	-0.030	0.013	-0.030	0.028	0.035	-
	(0.022)	(0.016)	(0.027)	(0.043)	(0.039)	
Age of Youngest Child (Annualized)	-0.000**	-0.004***	-0.006***	-0.010***	-0.008***	-
	(0.000)	(0.001)	(0.002)	(0.004)	(0.004)	
Sargan test	22.3(19)	34.8(19)**	30.0(19)*	32.2(31)	7.45(5)	175.0(149)*
m2 statistic	0.545	1.552	2.458**	-0.022	-0.457	0.219
Number of lags as instruments	12	12	12	18	5	5
Sample period (RA as month 0)	-23 to -4	0-23	4-53	4-23	4-23	4-23
Persons	2,570	2,570	2,570	2,570	2,570	2,570
Observations	43,690	53,970	120,790	43,690	43,690	43,690

1) The benchmark model is result (1) in Table 2.
2) "RA" denotes the time of random assignment. Annualized numbers are defined as original, monthly level coefficient estimates (standard errors) times 12. The first three months in the sample period are not directly used due to lagged dependent variables.
3) Results (1) to (5) reports two-step GMM estimates using the AH-type estimator developed in the estimation section. Result (6) reports two-step GMM estimates based on the Arellano-Bond (1991) estimator with instruments interacted with the FTP dummy. The degrees of freedom in the Sargan test of overidentifying restrictions are in parentheses. The m2 statistic refers to the Arellano-Bond (1991) specification test of second-order serial correlation in errors.
4) Asymptotic standard errors are in parentheses. *, **, *** denote significance at the 10, 5, 1 percent levels respectively.
5) In results (1) to (5), variables not reported include race, age, years of schooling, number of children, youngest child under 3 dummy, FTP dummy*youngest child under 3 dummy, 36-month time limit dummy, year dummies, and a constant. In result (6), variables not reported include year dummies and a constant.



(a) Welfare Participation Rate by Month



(b) Distribution of Total Months of Welfare Receipt since Random Assignment in Month 24 from Random Assignment

Figure 1: Welfare Participation Rate and Total Months of Welfare Receipt of FTP Control and Program Groups

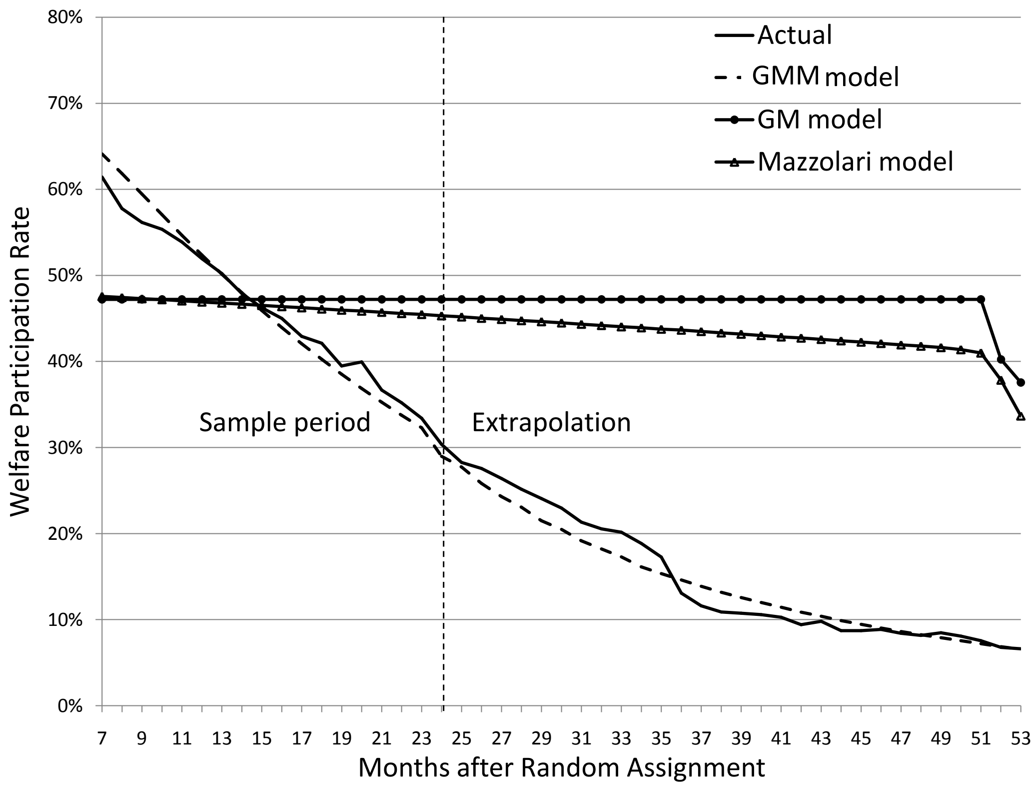
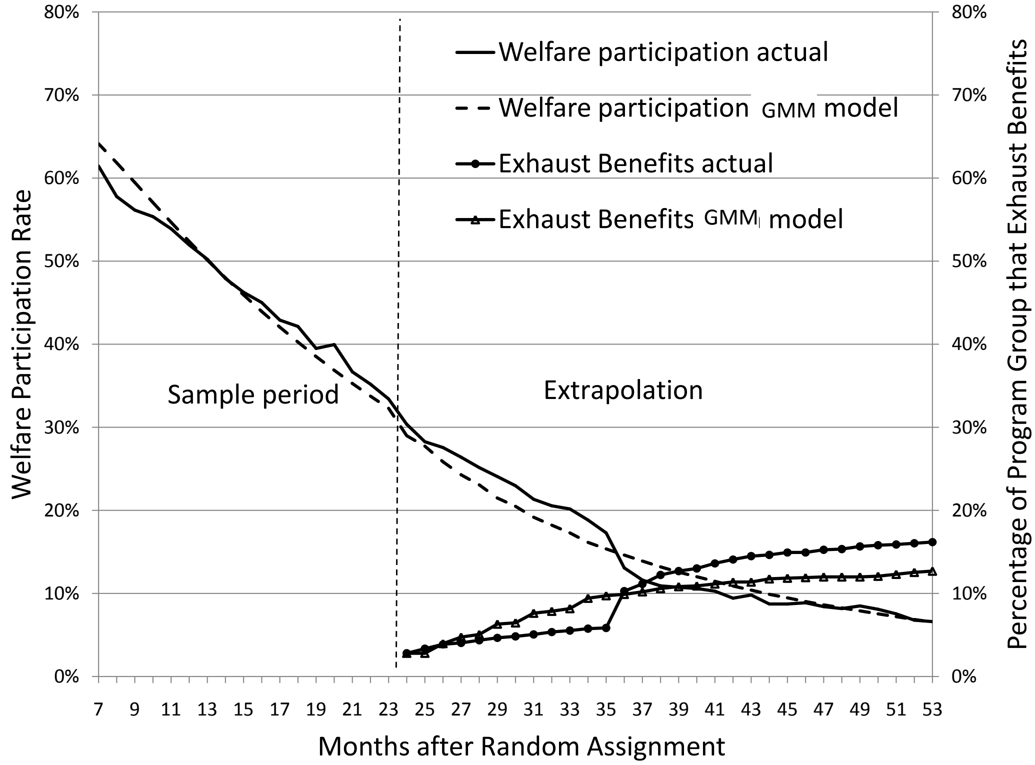
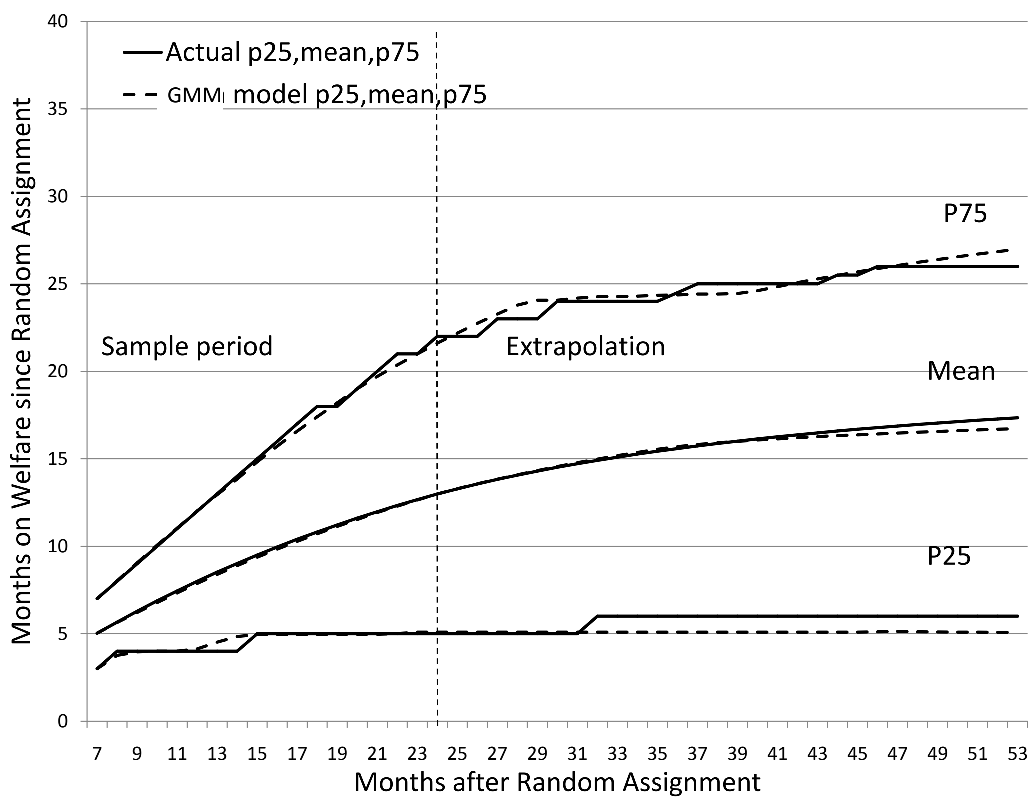


Figure 2: Predictive Performance of GM, Mazzolari and Chan Models



(a) Welfare Participation Rate and Benefit Exhaustion Rate



(b) 25th Percentile, Mean, and 75th Percentile of the Total Months of Welfare Receipt under the Time Limit

Figure 3: Within-Sample Prediction and Extrapolation of the Benchmark Model for the Program Group

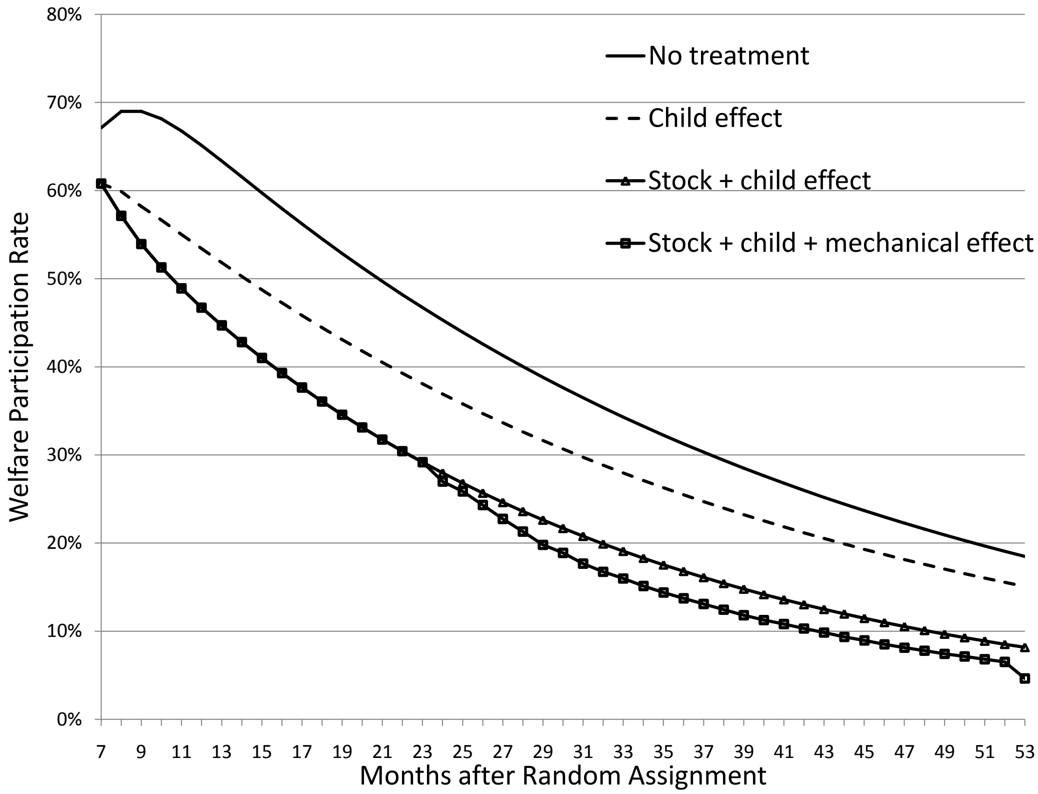
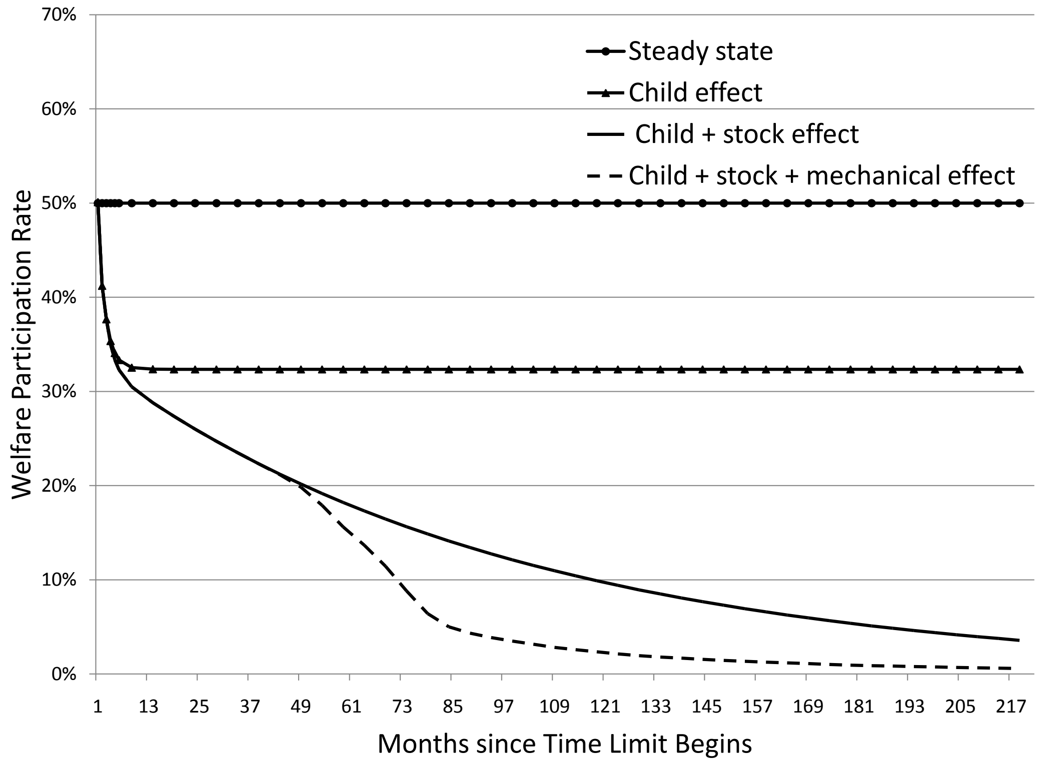
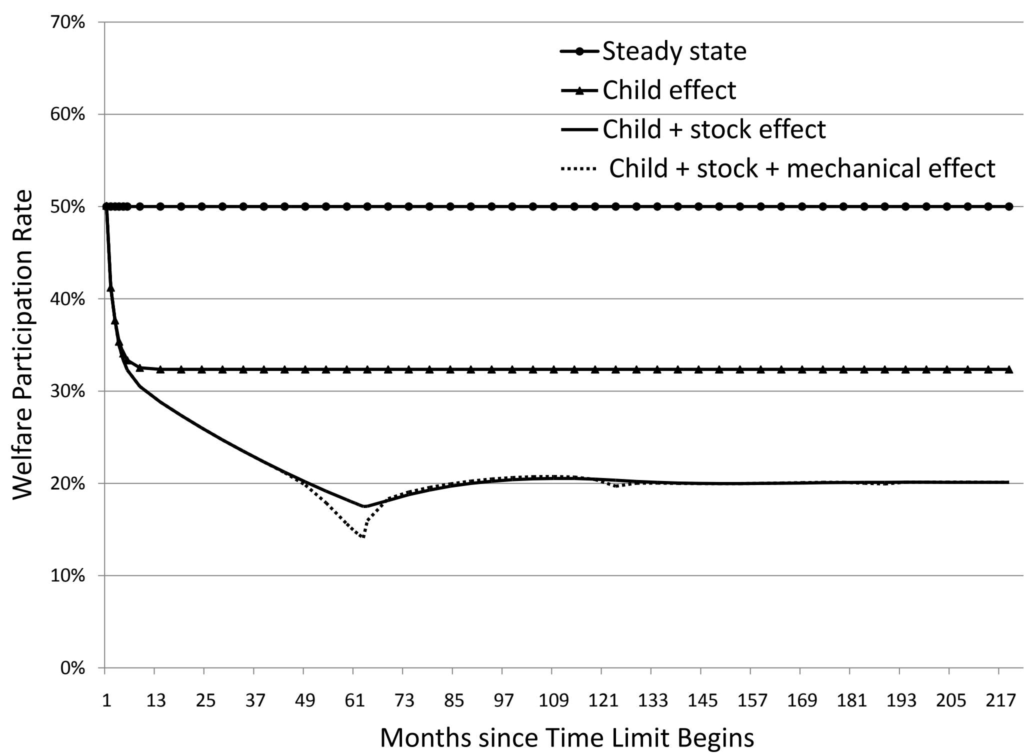


Figure 4: Treatment Effect of the Time Limit for the Program Group

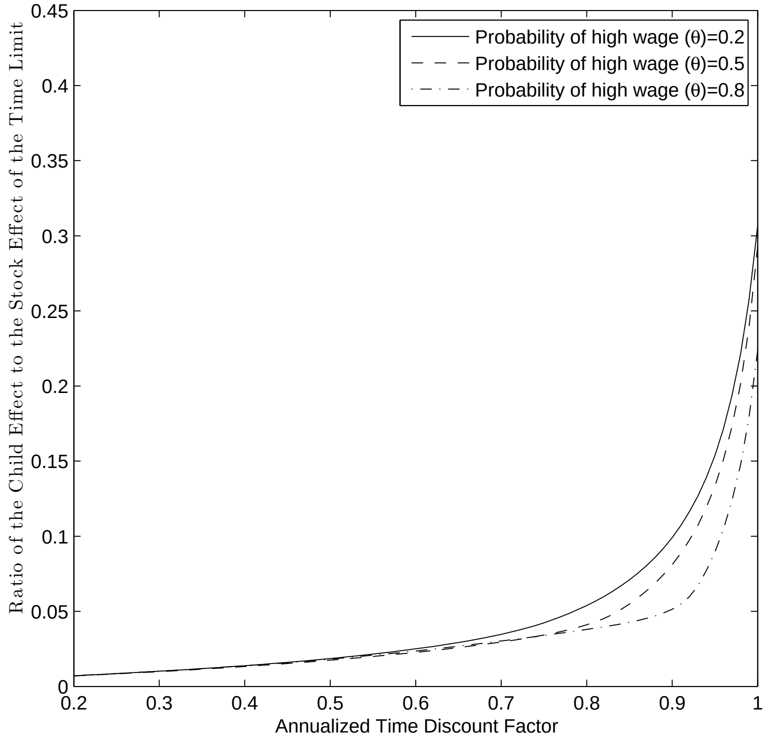


(a) A Permanent (Lifetime) Limit of 24 Months

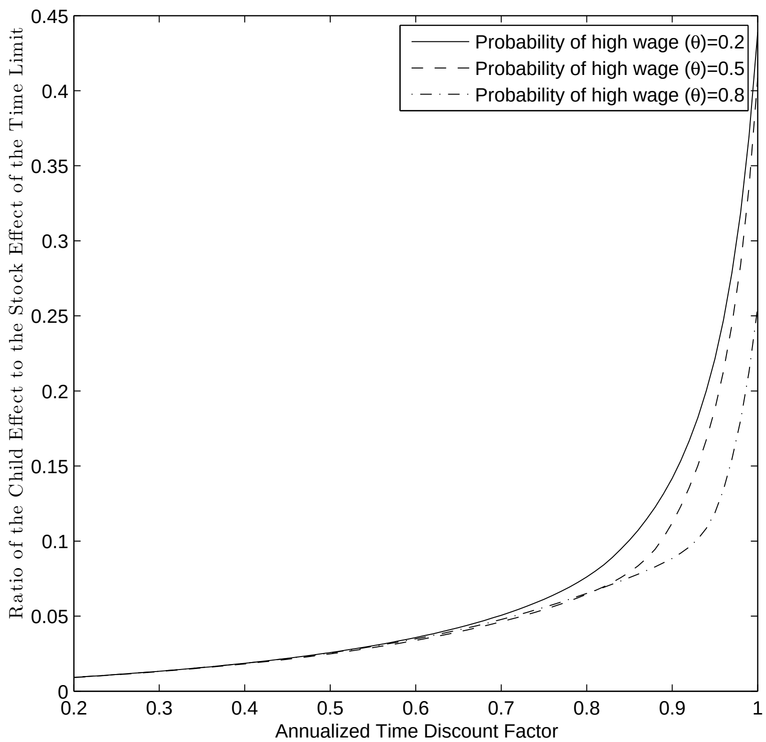


(b) A Maximum of 24 Months of Welfare Receipt in any 60-Month Period

Figure 5: The Effects of Imposing a Time Limit



(a) A Permanent (Lifetime) Limit of 24 Months



(b) A Permanent (Lifetime) Limit of 36 Months

Figure 6: Time Discount Factor and the Relative Behavioral Effects of a Time Limit

References

Ahn, S.C., and P. Schmidt (1995): "Efficient estimation of models for dynamic panel data," *Journal of Econometrics*, 68, 5–28.

Ai, Chunrong, and Edward Norton (2003): "Interaction Terms in Logit and Probit Models," *Economic Letters*, 80, 123–129.

Anderson, T.W., and C. Hsiao (1981a): "Estimation of Dynamic Models with Error Components," *Journal of the American Statistical Association*, 76, 598–606.

——— (1981b): "Formulation and Estimation of Dynamic Models using Panel Data," *Journal of Econometrics*, 18, 47–82.

Andren, Thomas (2007): "The Persistance of Welfare Participation," Discussion Paper No. 3100, IZA.

Arellano, Manuel, and Stephen Bond (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277–297.

Bitler, Marianne P., Jonah B. Gelbach, and Hilary W. Hoynes (2006): "What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments," *American Economic Review*, 96, 988–1012.

Bloom, Dan, and James Kemple et al. (2000): "The Family Transition Program: Final Report on Florida's Initial Time-Limited Welfare Program," Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families; and New York: MDRC.

Borsch-Supan, Axel, and Vassilis A. Hajivassiliou (1993): "Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models," *Journal of Econometrics*, 58, 347–368.

Bowsher, Clive G. (2002): "On Testing Overidentifying Restrictions in Dynamic Panel Data Models," *Economics Letters*, 77, 211–220.

Bruce Wallace and Tim Richards (2008): "The Rise and Fall of Welfare Time Limits in British Columbia," Victoria: Vancouver Island Public Interest Research Group.

Butler, J.S., and Robert Moffitt (1982): "A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model," *Econometrica*, 50, 761–764.

Card, David, and Dean Hyslop (2005): "Estimating the Effects of a Time-Limited Earnings Subsidy for Welfare-Leavers," *Econometrica*, 73, 1723–1770.

Chamberlain, G. (1985): "Heterogeneity, Omitted Variables Bias, and Duration Dependence," in *Longitudinal Analysis of Labor Market Data*, ed. by J.J. Heckman, and B. Singer. Cambridge University Press.

Chay, Kenneth, Hilary Hoynes, and Dean Hyslop (2001): "A Non-Experimental Analysis of True State Dependence in Monthly Welfare Participation Sequences," Working Paper.

Chay, Kenneth, and Dean Hyslop (2000): "Identification and Estimation of Dynamic Binary Response Panel Data Models: Empirical Evidence using Alternative Approaches," Working Paper.

Cox, D.R. (1958): "The Regression Analysis of Binary Sequences," *Journal of the Royal Statistical Society*, Series B, 20, 215–232.

Fang, Hanming, and Michael Keane (2004): "Assessing the Impact of Welfare Reform on Single Mothers," *Brookings Papers on Economic Activity*, (1), 1–116.

Grogger, Jeffrey (2003): "The Effects of Time Limits, the EITC, and Other Policy Changes on Welfare Use, Work, and Income Among Female-Headed Families," *Review of Economics and Statistics*, 85, 394–408.

——— (2004): "Time Limits and Welfare Use," *Journal of Human Resources*, 39, 405–424.

Grogger, Jeffrey, and Charles Michalopoulos (1999): "Welfare Dynamics Under Time Limits," NBER Working Paper.

——— (2003): "Welfare Dynamics Under Time Limits," *Journal of Political Economy*, 111, 530–554.

Ham, John, and Robert LaLonde (1996): "The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Training," *Econometrica*, 64, 175–205.

Hansen, Jorgen, Magnus Lofstrom, and Xuilin Zhang (2006): "State Dependence in Canadian Welfare Participation," Discussion Paper No. 2266, IZA.

Hansen, L.P. (1982): "Large Sample Properties of Generalized Mehod of Moments Estimators," *Econometrica*, 50, 1029–1054.

Hari, Johann (2008): "Beware the Tories' Wisconsin Welfare Plan," The Independent.

Heckman, James (1978): "Simple Statistical Models for Discrete Panel Data Developed and Applied to Test the Hypothesis of True State Dependence Against the Hypothesis of Spurious State Dependence," *Annales de l'INSEE*, 30-31, 227–269.

——— (1981a): "Heterogeneity and State Dependence," in *Studies in Labor Markets*, ed. by S. Rosen. University of Chicago Press.

——— (1981b): "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process and Some Monte Carlo Evidence," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by Charles Manski, and Daniel McFadden. Cambridge, Mass.: MIT Press.

——— (1981c): "Statistical Models for Discrete Panel Data," in *Structural Analysis of Discrete Data*, ed. by Charles Manski, and Daniel McFadden, p. Ch. 3. MIT Press.

Hendra, Rick (2002): "Florida's Family Transition Program Public Use File Data Documentation," Manpower Demonstration Research Corporation.

Horowitz, Joel, and Charles Manski (2000): "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data," *Journal of the American Statistical Association*, 95, 77–84.

Hyslop, Dean (1999): "State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women," *Econometrica*, 67, 1255–1294.

Karvelas, Patricia (2008): "Tony Abbott for 6-month time limit on dole," The Australian.

Keane, Michael (1993): "Simulation Estimation for Panel Data Models with Limited Dependent Variables," in *Handbook of Statistics*, ed. by C.R. Rao G.S. Maddala, and H.D. Vinod, p. Ch. 20. Amsterdam: Elsevier Science Publishers B.V.

——— (1994): "A Computationally Practical Simulation Estimator for Panel Data," *Econometrica*, 62, 95–116.

Keane, Michael, and Kenneth Wolpin (1997): "The Career Decisions of Young Man," *Journal of Political Economy*, 105(3), 473–522.

——— (forthcoming): "The Role of Labor and Marriage Markets, Preference Heterogeneity and the Welfare System in the Life Cycle Decisions of Black, Hispanic and White Women," *International Economic Review*.

Lee, David (2009): "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *Review of Economic Studies*, 76, 1071–1102.

Mazzolari, Francesca (2007): "Welfare Use when Approaching the Time Limit," *Journal of Human Resources*, 42, 596–618.

Mazzolari, Francesca, and Giuseppe Ragusa (2009): "Time Limits: The Effects on Welfare Use and other Consumption-Smoothing Mechanisms," Working Paper.

Moffitt, Robert (1999): "The Effect of Pre-PRWORA Waivers on AFDC Caseloads and Female Earnings, Income, and Labor Force Behavior," in *Economic Conditions and Welfare Reform*, ed. by Sheldon H. Danziger. Kalamazoo: W.E. Upjohn Institute for Employment Research.

Plant, M.W. (1984): "An Empirical Analysis of Welfare Dependence," *American Economic Review*, 74, 673–694.

Rich, Katherine (2003): "Saving the Next Generation from Welfare Dependency," New Zealand National Party Discussion Paper.

Sargan, J.D. (1958): "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica*, 26, 393–415.

Schochet, Peter, John Burghardt, and Sheena McConnell (2008): "Does Job Corps Work? Impact Findings from the National Job Corps Study," *American Economic Review*, 98, 1864–1886.

Schoeni, Robert, and Rebecca Blank (2000): "What has Welfare Reform Accomplished? Impacts on Welfare Participation, Employment, Income, Poverty, and Family Structure," Labor and Population Program Working Paper 00-02. Santa Monica: RAND.

Swann, Christopher A. (2005): "Welfare Reform when Recipients are Forward-Looking," *Journal of Human Resources*, 40, 31–56.

U.S. Council of Economic Advisors (1997): "Explaining the Decline in Welfare Receipt 1993-1996," Washington, DC.

——— (1999): "The Effects of Welfare Policy and the Economic Expansion on Welfare Caseloads: An Update," Washington, DC.

Ziliak, James P., David N. Figlio, Elizabeth E. Davis, and Laura S. Connolly (2000): "Accounting for the Decline in AFDC Caseloads: Welfare Reform or the Economy?," *Journal of Human Resources*, 35, 570–586.