UNIVERSITY OF TECHNOLOGY SYDNEY

ECONOMICS DISCIPLINE GROUP

UTS BUSINESS SCHOOL

WORKING PAPER NO. 23 August 2014

Measuring the Dynamic Effects of Welfare Time Limits

Marc K. Chan

ISSN: 2200-6788

http://www.business.uts.edu.au/economics/

Measuring the Dynamic Effects of Welfare Time Limits

Marc K. Chan *
University of Technology Sydney

August 13, 2014

Abstract

This paper develops a new dynamic panel data model that can formally incorporate the dynamics of welfare participation behavior under time limits. The model is estimated using data from a policy experiment in the United States and a generalized method of moments estimator. The effects of time limits are found to be larger and more dynamical than in previous regression approaches, after comparing estimation results from several approaches using the same data. Around 40 percent of the anticipatory effect of the time limit are due to individuals depleting their stock of remaining months of welfare eligibility. The anticipatory effect is also found to be much larger among disadvantaged individuals. The results call for further assessment of the importance of time limits in explaining the low welfare caseload in the post-welfare reform era.

JEL CLASSIFICATION: I38, C23

Keywords: Welfare time limit, dynamic panel data model, generalized method of moments, policy experiment

*I am grateful to Robert Moffitt, Jeffrey Grogger, and participants of the Society of Labor Economists and Econometric Society Meetings for their comments. 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. Address: Economics Discipline Group, University of Technology Sydney, PO Box 123, Broadway, NSW 2007, Australia. Email: marc.chan@uts.edu.au.

1 Introduction

The welfare system in the United States experienced dramatic changes in the 1990s. Of all the policy changes that occurred during the reform period, welfare time limits were arguably the most controversial and radical. Prior to the passage of the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) in 1996, female heads of family who were in poverty and had at least one child under 18 years of age were entitled to cash benefits under the Aid to Families with Dependent Children (AFDC) program. By contrast, the Temporary Aid to Needy Families (TANF) program, which replaced AFDC under PRWORA, restricted families to a maximum of five years of federally funded benefits. Many states imposed even more stringent limits.

Following a steep decline in the national welfare caseload in the late 1990s, it remains controversial as to why welfare caseloads remained at a low level throughout the 2000s.[1] Since welfare time limits represent a very substantial reduction in the generosity of welfare benefits – for a single mother with a newborn infant, the de facto reduction in welfare benefit is close to 75 percent under the federal limit – conventional wisdom suggests that the policy should have a sizable effect on welfare participation in terms of behavioral response or mechanical termination of benefits. In this respect, the size of the effect is very much a key research question in the large literature on the incentive effects of the welfare system (e.g., Moffitt (1992)). At a policy level, this is also an important issue because the number of terminated cases is heavily affected by the behavioral incentive to reduce welfare use prior to reaching the time limit.

However, due to the complicated behavioral dynamics, the literature has yet developed a methodological framework that can fully encapsulate the dynamic mechanism generated by time limits, while at the same time avoid restrictive assumptions in model specification. A variety of approaches has been developed to measure the behavioral effects of time limits on welfare participation. Earlier studies such as Ziliak et al. (2000) and Blank (2001) analyze the effect of AFDC waiver programs, including those that involve time limits, on welfare caseloads. Grogger and Michalopoulos (2003) use the theoretical prediction that the time limit should result in a larger reduction in welfare use among families with younger children, and they exploit the quasi-experimental variation in the age of the youngest child in the family to isolate the effects of time limits from other

[1] There has been limited study on how various factors contributed to the low caseloads in recent years. One of the exceptions is Bitler and Hoynes (2010), who analyzed the cyclicality of welfare caseloads in the 2007-09 recession.

policies.[2] Fang and Keane (2004) and Mazzolari (2007) extend upon Grogger and Michalopoulos (2003) by also attempting to measure the reduction in welfare use when individuals deplete their stock of remaining periods of welfare eligibility. In particular, Mazzolari (2007) uses differences in timing and stringency of state time limit policies, as well as pre-time-limit welfare participation rates by sociodemographic characteristics to construct a prediction of the stock of remaining periods of welfare eligibility at each point in time.[3] Ribar et al. (2008) use a competing-risks model and administrative data from the post-reform period to analyze the effects of food stamp recertification and time limits in South Carolina. Although they do not use the quasi-experimental variation in the age of the youngest child in the family, they find that welfare recipients tend to exit welfare before reaching the time limit. In dynamic structural models such as Swann (2005), Keane and Wolpin (2010) and Chan (2013), the stock of remaining periods of welfare eligibility enters as a state variable in the dynamic optimization problem, but a whole set of utility structure and budget constraints has to be formally specified.

This paper develops and estimates a new dynamic panel data model that can formally incorporate welfare dynamics under time limits. It is related to existing regression models, in particular, Grogger and Michalopoulos (2003) and Mazzolari (2007), and is distinct from the above work in the treatment of dynamics and unobserved heterogeneity, as well as the utilization of the panel feature of individual-level data for identification. While the model formally incorporates dynamics, it avoids imposing behavioral assumptions that are common in structural models. The model is estimated using a generalized method of moments (GMM) estimator and data from a randomized welfare reform experiment in the 1990s. In the experiment, only the treatment group was subject to a time limit. To take into account of the endogeneity of the stock of remaining periods of welfare eligibility, instruments are constructed in a similar spirit to the literature of dynamic panel data models (e.g., Anderson and Hsiao (1982) and Arellano and Bond (1991)).

The analysis makes two contributions related to the application of dynamic panel data models in policy evaluation. First, it shows how additional moment conditions can be used for identification when a policy introduces additional dynamics to behavior. Since the dynamics are fundamentally

[2] A similar approach is pursued by Grogger (2003) and Grogger (2004).

[3] The Mazzolari model is estimated using the method of two-stage least squares, where the first stage uses functions of the predicted stock as instruments. Fang and Keane (2004) exploit the differences in timing and stringency of state time limit policies, but they use the time elapsed since the implementation of the time limit policy to construct a prediction of the stock of remaining periods of welfare eligibility.

different in the treatment group due to the presence of a time limit, the set of instruments has to be customized accordingly for model estimation. Second, it shows that the endogeneity of work-related variables (if included as regressors) can be resolved using additional instruments. This approach avoids the joint modeling of welfare participation and employment decisions, which can be advantageous in a number of situations.

This paper also brings together experimental policy evaluation with the growing reduced-form literature that uses panel data to study welfare dynamics. It is related to Klerman and Haider (2004), who develop a stock-flow model to explain the dynamics of welfare caseloads in California. In their model, the "stock" of welfare caseloads, which represents the total *number* of welfare recipients in a region, consists of individuals who have been welfare recipients for different numbers of periods. They show that a relatively simple flow model can generate rich dynamics in the stock of welfare caseloads. While their definition of the "stock" is fundamentally different from ours, our model is similar in spirit to their flow model. In particular, our model is richer in terms of incorporating individual-level unobserved heterogeneity, and allowing for both policy and socioeconomic factors to influence individual behavior.[4]

The estimation results from our model suggest that the effects of welfare time limits are larger and more dynamical than in previous regression approaches. In the baseline sample, the overall anticipatory effect of the time limit becomes twice as large just prior to two years after random assignment than one year after random assignment. Around 40 percent of the overall anticipatory effect of the time limit are due to the depletion of the stock of remaining welfare eligibility, and the rest are due to the Grogger and Michalopoulos (2003) effect. In addition, all else being equal, welfare participation decreases by 7.2 percentage points for each additional cumulative year of welfare use under the time limit. Somewhat surprisingly, we find that the anticipatory effect of the time limit is larger among disadvantaged individuals due to their higher dependence on welfare.

The paper also shows analytically and empirically that a dynamic model specification is very important for the identification and measurement of the effects of time limits. Using the same data, we compare the estimation results between our model and existing quasi-experimental and instrumental variable approaches. In the baseline sample, the existing approaches tend to underestimate

[4]The econometric analysis in Klerman and Haider (2004) is also different from ours as it is based on the small N, large T framework.

the size of the stock effect and overestimate the size of the Grogger and Michalopoulos (2003) effect. In addition, among disadvantaged individuals, both approaches tend to underestimate the overall effect of the time limit by approximately 40 percent. Although the quantitative findings are based on a policy experiment, the results have interesting implications on the overall role of time limits in at least two aspects. First, time limits should have played a larger role in explaining the caseload decline in the 1990s and the low level of caseloads in the 2000s. A strong effect appears to be more in line with the large reduction in the generosity of welfare benefits due to time limits. Second, a large behavioral effect implies that fewer families will actually exhaust their benefits under the time limit. This may have played a role in explaining the relatively low number of terminated cases due to time limits in recent years.

This paper is organized as follows. Section 2 describes the policy experiment and the data used for the analysis. Section 3 presents our baseline model of time limits, and compares its properties with existing regression approaches, in particular, Grogger and Michalopoulos (2003) and Mazzolari (2007). Section 4 discusses the GMM estimator and instruments for the baseline model. Sections 5 and 6 report estimation results from the baseline sample and the disadvantaged sample, respectively. In both sections, comparison will be made with existing regression approaches. Section 7 concludes.

2 Policy Background and Data

The empirical analysis is based on Family Transition Program (FTP), which was one of the few welfare reform experiments in the United States to impose a welfare time limit. FTP was a pilot program that operated from 1994 to 1999 in Escambia, a mid-sized county in northwest Florida. From May 1994, the program began to assign welfare applicants and recipients randomly into the control group (AFDC group) or the treatment group (FTP group). Individuals in the control group were subject to the preexisting rules of Aid to Families with Dependent Children program (AFDC), which had no welfare time limits. The treatment group was subject to a time limit, which restricted families to a maximum number of months of cash assistance within a given period. The treatment group was also subject to financial work incentives, in which more earnings were disregarded in the calculation of welfare benefits, as well as enhanced employment services.[5] There are several

[5]In the treatment group, the first 200 dollars of earnings were disregarded, and all earnings in excess of 200 dollars were subject to a benefit reduction rate of 50 percent. In the control group, the default earnings disregard was 120

features of the time limit that allow us to disentangle its effects from other aspects of the FTP treatment. The features and the identification strategy will be described in detail in the model section.

At the time of FTP implementation, the socioeconomic environment in Escambia county was largely in line with the overall environment in Florida. Relative to the state, the county had a slightly lower level of median household income ($25,158) and unemployment rate (4.5 percent), and a larger proportion of the nonwhite population (23.4 percent) (Bloom et al. (2000)). As a whole, the state of Florida was similar to the United States in the above characteristics. In 1994, Florida constituted 5.4 percent of the population in the United States. Approximately 250 thousand families received cash assistance, which represented 5.5 percent of the nationwide caseload.

The source of the data is the FTP public use file, which consists of individuals who were randomly assigned by the program between May 1994 and February 1995. The analysis focuses on individual behavior within the first 24 months following random assignment, so none of the individuals will have reached the time limit by the end of the sample period. The data contains demographic information collected at the time of random assignment, and administrative record on monthly welfare participation. Individuals are classified into one of the two subsamples depending on her level of disadvantage, which is defined by several observable characteristics.[6] If a treatment group individual is not considered as disadvantaged, she will be restricted to a maximum of 24 month of cash assistance within a 60-month period; otherwise, she will be restricted to a maximum of 36 months of cash assistance within a 72-month period. Control group individuals are subject to the same AFDC rules regardless of the level of disadvantage. The nondisadvantaged individuals constitute the *baseline sample*, which contains 1306 individuals (589 in the control group and 717 in the treatment group). The disadvantaged individuals constitute the *disadvantaged sample*, which contains 1281 individuals (698 in the control group and 583 in the treatment group).

The Family Transition Program is chosen for analysis due to its relatively early launch compared to the implementation of the state's TANF program at the end of 1996. In this respect, FTP arguably represents one of the best policy environments for studying time limits.[7] Between 1994

dollars with a benefit reduction rate of 67 percent for earnings in excess of 120 dollars.

[6]An individual is considered as disadvantaged if she had received AFDC for at least 36 of the 60 months prior to random assignment, or was under 24 years old and had no high school diploma and no work experience one year prior to random assignment. For more information, see Bloom et al. (2000).

[7]Another welfare reform experiment that has been extensively studied is Connecticut's Jobs First Program, which

and 1996, the welfare caseload in Florida was subject to a mild reduction at a rate of 10 percent annually. In the absence of other major policy changes, this provides a time window in which the effects of FTP components can be properly measured (Bloom et al. (2000)). Our analysis also allows for a more direct comparison with Grogger and Michalopoulos (2003), who use data from FTP for analysis. However, our sample is different, as their data also includes individuals who were randomly assigned between March 1995 and October 1996.

Table I reports summary statistics of the baseline and disadvantaged samples. Individuals in the baseline sample are less likely to be black, receive more education, and have less children in the family than individuals in the disadvantaged sample. In the baseline sample, the average welfare participation rate during the sample period is 37.55 percent in the control group and 36.06 percent in the treatment group. By contrast, in the disadvantaged sample, the average welfare participation rate is 59.42 percent in the control group and 66.36 percent in the treatment group. This suggests that the program has a differential impact on the baseline and disadvantaged samples. In particular, the FTP tends to be more attractive to disadvantaged individuals. Both samples are subject to a decline in the welfare participation rate over time. Moreover, there is a tendency for the welfare participation rate in the treatment group to drop faster towards the end of the sample period.

The bottom panel of the table reports the welfare participation rates conditional on cumulative months of welfare use since random assignment.[8] The results suggest that the time limit may have played a role in reducing welfare use, especially among individuals who are close to reaching their time limit. For instance, in the baseline sample, there is only a small control-treatment difference in the welfare participation rate among individuals who have used few months of welfare. By contrast, among individuals who have used more months of welfare, the welfare participation rate tends to be significantly lower in the treatment group than the control group. In the disadvantaged sample, the welfare participation rate is higher in the treatment group than the control group among individuals who have used few months of welfare. However, among individuals who have used more months of welfare, the welfare participation rate tends to become similar between the control and treatment groups.

---

started in 1996. One of the features of Jobs First is a 21-month time limit. Existing studies on the program, such as Bitler et al. (2006) and Kline and Tartari (2014), have focused on the effects of financial work incentives (i.e., earnings disregards), but do not analyze the time limit.

[8]Due to heterogeneity, the welfare participation rate tends to be higher among individuals who have used more months of welfare.

3 Baseline Model and Implications for Typical Regression Approaches

A dynamic panel data model of time limits is first presented below. In the baseline sample, the baseline version of the model is:

$$\begin{aligned} y_{it} = \lambda_0 + \sum_{k=1}^{K} \alpha_k y_{i,t-k} + \beta_0 E_i \mathbf{1}\{S_{it} < H_{it}\} + \beta_S E_i \mathbf{1}\{S_{it} < H_{it}\}(S_{it} - 24) + \\ \beta_A E_i \mathbf{1}\{S_{it} < H_{it}\}(A_{it} - 16) + \beta_E E_i + \boldsymbol{X_{it}} \boldsymbol{\lambda} + \mu_i + \epsilon_{it}. \end{aligned} \tag{1}$$

In the model for the disadvantaged sample, the terms $S_{it} - 24$ and $A_{it} - 16$ are replaced by $S_{it} - 36$ and $A_{it} - 15$, respectively. The month in which random assignment occurs is defined as month 0. For individual i at month t, y_{it} represents an indicator of welfare program participation, E_i is the treatment group indicator, and S_{it} is the "stock" of remaining months of welfare eligibility for a treatment group individual under the time limit:

$$S_{it} = S_{i0} - \sum_{k=0}^{t-1} y_{ik}. \tag{2}$$

At month 0, all treatment group individuals in the baseline sample have 24 remaining months of welfare eligibility, that is, $S_{i0} = 24$. In the disadvantaged sample, the initial stock is $S_{i0} = 36$ instead. The stock S_{it} becomes one unit lower next period when the individual participates in welfare this period, and remains unchanged otherwise. When S_{it} reaches zero, the individual exhausts all her months of welfare eligibility, and she becomes ineligible for welfare.

The variable H_{it} denotes the number of remaining months until the individual's youngest child becomes 18 years of age. It captures the length of the remaining time horizon in which the individual is potentially eligible for welfare (Grogger and Michalopoulos (2003)). The variable can be defined as $H_{it} = 12 \times (18 - A_{it})$, where A_{it} denotes the age of the youngest child (in years). The vector $\boldsymbol{X_{it}}$ contains a set of demographic characteristics including the individual's age, race (equal to one if she is black and zero otherwise), years of schooling, the number of children in the family, the age of the youngest child in the family, a dummy variable that equals one if the youngest child is under age 3 and zero otherwise, and the interaction between the dummy variable above with the treatment group indicator.[9] It also contains calendar year indicators, calendar month indicators to capture seasonality, and a quartic time trend.[10] The unobserved effect is denoted by μ_i, which represents individual-level unobserved heterogeneity. Transitory errors of the model are denoted by ϵ_{it}, which are uncorrelated with the unobserved effect and are serially uncorrelated, an assumption which we will return later. For the individual, the decision to participate in welfare in the current period depends on welfare participation in K earlier periods (i.e., $y_{i,t-1}, ..., y_{i,t-K}$). The corresponding state dependence parameters are denoted by α_k, where $k = 1, ..., K$.

In the above dynamic specification, the effect of the time limit can be disentangled from other aspects of FTP through three channels. The first key observation for identification is that the time limit policy introduces additional dynamics to the behavior of selected treatment group individuals. In particular, the time limit generates a behavioral effect only when the stock of remaining months of welfare eligibility is *lower* than the remaining time horizon, that is, $\mathbf{1}\{S_{it} < H_{it}\} = 1$, where $\mathbf{1}\{.\}$ is an indicator function. For other treatment group individuals, the time limit is not binding and therefore it does not have any effect on behavior. Since these individuals are still subject to other aspects of the FTP treatment, their levels of welfare participation are different from the control group – the effect of other aspects of FTP can therefore be isolated out in the parameter β_E (Grogger and Michalopoulos (2003)). We expect β_E to be positive due to more generous financial work incentives in the treatment group, but the overall sign could be ambiguous if individuals place substantial disutility on enhanced employment services.[11]

In the baseline sample and model, the parameter β_0 is used for capturing the effect of the time limit for a treatment group individual who has 24 remaining months of welfare eligibility and whose youngest child is 16 years of age. In the disadvantaged sample, the corresponding thresholds are 36 months and 15 years, respectively. For the above types of individuals, the time limit is irrelevant, so we should expect β_0 to be zero.

The next two channels characterize the specific behavioral dynamics that the time limit induces

[9] The interaction variable captures the difference in exemption from mandatory work-related activities between control and treatment group individuals who have a child who is under 3 years old (Grogger and Michalopoulos (2003)).

[10] The quartic time trend is expressed in monthly units. It does not contain a linear term, which is collinear with calendar year and calendar month indicators.

[11] In the baseline model, other aspects of FTP can generate lagged effects on welfare participation via autoregressive parameters α_k. In an extended model, we also explicitly control for work-related variables to account for potentially different levels of accumulated work experience between the control and treatment groups. See the subsection on additional specifications for more details.

among treatment group individuals. First, all else being equal, a treatment group individual should reduce welfare participation further when she has a lower stock of remaining months of welfare eligibility under a time limit. This "stock" effect is captured by the parameter β_S, which has positive expected sign. This implies that the welfare participation rate of treatment group individuals will be a direct function of the remaining stock S_{it}; by contrast, all else being equal, the "stock measure" will not affect the welfare participation behavior of control group individuals.

Second, the time limit should reduce welfare participation by a larger degree, the longer the remaining time horizon. In the literature this is called the *GM effect* (Grogger and Michalopoulos (2003)), and it is captured by the parameter β_A, which has positive expected sign.

In the discussion below, the baseline model will be compared with two existing regression approaches on time limits. It will be shown that our model has several advantages via incorporating dynamics formally and utilizing individual-level information in the panel data for identification.

Grogger and Michalopoulos (2003). The *GM model* uses the quasi-experimental variation in the age of the youngest child of the family to identify the effects of time limits. This approach is widely used in subsequent studies. In the baseline sample, the model that most closely resembles the original specification is:

$$y_{it} = \lambda_0 + \beta_A E_i \mathbf{1}\{A_{it} < 16\}(A_{it} - 16) + \beta_E E_i + \boldsymbol{X_{it}} \boldsymbol{\lambda} + \epsilon_{it}. \tag{3}$$

Their model is estimated using ordinary least squares (OLS) on individual-level data from the first 24 months following random assignment.[12]

We will show analytically that controlling for the stock S_{it} is important in obtaining an unbiased estimate of the GM effect β_A. To illustrate the argument, consider the following data generating process, which is a simplified version of equation (1):

$$y_{it} = \lambda_0 + \beta_S E_i(S_{it} - 24) + \beta_A E_i(A_i - 16) + \lambda_A A_i + \epsilon_{it}, \tag{4}$$

[12]Their model uses Huber-White standard errors to account for groupwise dependence in panel data. Their original specification is different in the following ways. First, welfare use and employment during the pre-random-assignment period are included as regressors, but calendar month indicators and a time trend are not included. Second, both the baseline and disadvantaged samples are pooled together for estimation, and a 36-month limit indicator variable is included as a regressor. Third, children under three years of age are handled in a slightly different way.

where $A_i < 16$ is the age of the youngest child of individual i. Rearranging the terms, we have

$$y_{it} = \lambda_0^{'} + \beta_S E_i S_{it} + \beta_A E_i A_i + \lambda_A A_i + \epsilon_{it}, \tag{5}$$

where $\lambda_0^{'} \equiv \lambda_0 - (24\beta_S + 16\beta_A)E_i$. Now consider a cross-section of the data at month $t > 0$, and two values of $A_i = a, a^{'}$, where $a^{'} > a$. Then, we have

$$\lambda_A(a^{'} - a) = E(y_{it}|A_i = a^{'}, E_i = 0) - E(y_{it}|A_i = a, E_i = 0), \tag{6}$$

$$(\lambda_A + \beta_A)(a^{'} - a) = E(y_{it}|A_i = a^{'}, E_i = 1, S_{it} = s) - E(y_{it}|A_i = a, E_i = 1, S_{it} = s), \tag{7}$$

which implies that the GM effect is

$$\begin{aligned}\beta_A = (a^{'} - a)^{-1}\Big[\Big(E(y_{it}|A_i = a^{'}, E_i = 1, S_{it} = s) - E(y_{it}|A_i = a, E_i = 1, S_{it} = s)\Big) - \\ \Big(E(y_{it}|A_i = a^{'}, E_i = 0) - E(y_{it}|A_i = a, E_i = 0)\Big)\Big].\end{aligned} \tag{8}$$

However, in the GM model, the GM effect is obtained *without* the conditioning on S_{it}:

$$\begin{aligned}\tilde{\beta}_A = (a^{'} - a)^{-1}\Big[\Big(E(y_{it}|A_i = a^{'}, E_i = 1) - E(y_{it}|A_i = a, E_i = 1)\Big) - \\ \Big(E(y_{it}|A_i = a^{'}, E_i = 0) - E(y_{it}|A_i = a, E_i = 0)\Big)\Big].\end{aligned} \tag{9}$$

Propositions 1 and 2 in the Appendix show that the GM effect obtained from the GM model ($\tilde{\beta}_A$) is generally different from the true GM effect (β_A):

$$\begin{aligned}\tilde{\beta}_A - \beta_A =& (a^{'} - a)^{-1}\beta_S \Big(E(S_{it}|A_i = a^{'}, E_i = 1) - E(S_{it}|A_i = a, E_i = 1)\Big) \\ &\begin{cases} = 0 & \text{if} \quad \beta_S = 0 \quad \text{or} \quad \lambda_A + \beta_A = 0, \\ > 0 & \text{if} \quad \beta_S > 0 \quad \text{and} \quad \lambda_A + \beta_A < 0, \\ < 0 & \text{if} \quad \beta_S > 0 \quad \text{and} \quad \lambda_A + \beta_A > 0. \end{cases}\end{aligned} \tag{10}$$

The intuition is described as follows. For simplicity, suppose welfare participation is a decreasing function of the age of the youngest child, that is, $\lambda_A < 0$. Moreover, suppose the stock effect is positive (i.e., $\beta_S > 0$), but the true GM effect is zero (i.e., $\beta_A = 0$). Then, as time progresses, treatment group individuals with a very young child will deplete their stock of remaining months of welfare eligibility *faster* than treatment group individuals who have a relatively old child. Since the former group tends to have a lower stock S_{it}, the time limit will reduce welfare participation by a larger degree in the former group. Therefore, when comparing the welfare participation behavior between both groups using the GM model, we will erroneously conclude that the GM effect is positive (i.e., $\tilde{\beta}_A > 0$), but the estimate is simply picking up part of the stock effect.

Equation (10) also implies that the overall effect of the time limit may be misrepresented as well. To illustrate the point, consider a special case where $\lambda_A = \beta_A = 0$ but $\beta_S > 0$. Then, under the time limit, treatment group individuals with a very young child will deplete their stock of remaining months of welfare eligibility *at the same rate* as treatment group individuals who have a relatively old child. As a result, the stock effect will reduce welfare participation equally in both groups. The GM model will correctly conclude that the GM effect is zero (i.e., $\tilde{\beta}_A = 0$), but it will fail to pick up the stock effect.[13]

A special case is that the GM model will generate an unbiased estimate of the GM effect β_A if data from month 0 is used only. However, the model will not be able to measure the stock effect, and will therefore underpredict the overall effect of the time limit in subsequent periods. More discussion will be provided in the Results section.

Mazzolari (2007). The Mazzolari model can be considered as an extension of the GM model. It uses an instrumental variable approach to remove the endogeneity of the stock S_{it}.[14] The baseline version that most closely resembles the original specification is:

$$y_{it} = \lambda_0 + \beta_0 E_i \mathbf{1}\{S_{it} < H_{it}\} + \beta_1 E_i \mathbf{1}\{S_{it} < H_{it}\} \frac{S_{it}}{H_{it}} + \beta_E E_i + \boldsymbol{X_{it}\lambda} + \epsilon_{it}. \tag{11}$$

The parameter β_0 captures the effect of the time limit when stock S_{it} approaches zero, and its expected sign is negative. The parameter β_1 captures the change in welfare participation as stock S_{it} becomes larger or the time horizon H_{it} becomes smaller. It reflects a combination of both the

[13] If the treatment group indicator E_i is included as a regressor, its coefficient β_E will pick up the average stock effect. However, this is not separately identifiable from other aspects of the FTP treatment.

[14] The stock S_{it} is endogenous because it is potentially correlated with the error term, which may contain individual-level unobserved heterogeneity.

GM and stock effects, and is expected to have a positive sign.[15] Mazzolari constructs the following variable as a proxy for S_{it}:

$$Z_{ijt} = S_{i0} - \tilde{k}_j t, \tag{12}$$

where $\tilde{k}_j$ is the average welfare participation rate in sociodemographic group j *before* the time limit policy was implemented, and Z_{ijt} is the simulated stock of remaining months of welfare eligibility that depletes linearly at a rate equal to $\tilde{k}_j$.[16] The model becomes similar to the GM model if $\tilde{k}_j = 0$, and it becomes similar in spirit to Fang and Keane (2004) if $\tilde{k}_j = 1$. Estimation is carried out using two-stage least squares (2SLS), where the first stage uses $\mathbf{1}\{Z_{ijt} < H_{it}\}$ and $\mathbf{1}\{Z_{ijt} < H_{it}\}\frac{Z_{ijt}}{H_{it}}$ as instruments for $\mathbf{1}\{S_{it} < H_{it}\}$ and $\mathbf{1}\{S_{it} < H_{it}\}\frac{S_{it}}{H_{it}}$.

Although the instruments remove the endogeneity of S_{it}, they also remove important individual-level information that can identify the effect of the time limit. To illustrate the argument, consider a simplified version of equation (11), where $H_{it} = \bar{H}$ is large enough such that $\mathbf{1}\{S_{it} < \bar{H}\} = 1$ for all i. Then, we have

$$y_{it} = \lambda_0 + \beta_0^{'} E_i + \beta_1^{'} E_i S_{it} + \boldsymbol{X_{it}\lambda} + \epsilon_{it}, \tag{13}$$

where $\beta_0^{'} = \beta_0 + \beta_E$ and $\beta_1^{'} = \beta_1/\bar{H}$. The two-stage least squares procedure replaces S_{it} with a linear projection on Z_{ijt}:

$$\begin{aligned} y_{it} &= \lambda_0 + \beta_0^{'} E_i + \beta_1^{'} E_i(\hat{\gamma}_0 + \hat{\gamma}_1 Z_{ijt}) + \epsilon_{it} \\ &= \lambda_0 + \beta_0^{'} E_i + \beta_1^{'} E_i(\hat{\gamma}_0 + \hat{\gamma}_1 (S_{i0} - \tilde{k}_j t)) + \epsilon_{it} \\ &= \lambda_0 + [\beta_0^{'} + \beta_1^{'}(\hat{\gamma}_0 + \hat{\gamma}_1 S_{i0})]E_i - \beta_1^{'}\hat{\gamma}_1 E_i \tilde{k}_j t + \epsilon_{it}. \end{aligned} \tag{14}$$

Suppose the stringency of the time limit is the same for all treatment group individuals (e.g., in the baseline sample, it is $S_{i0} = 24$). Then, according to equation (14), welfare participation under the

[15] The original specification is estimated using data from multiple periods in the Survey of Income and Program Participation (SIPP), and is different in the following ways. First, $\mathbf{1}\{S_{it} \leq 0\}$ is included as a regressor to capture the behavior of individuals who have reached the time limit during the sample period. Second, state-level regressors are included.

[16] Following Mazzolari, four sociodemographic groups are constructed using the individual's education (2 levels) and number of children in the family (2 levels). In estimation, the pre-time-limit welfare participation rates ($\tilde{k}_j$) are computed using data from 1 to 12 months prior to random assignment.

time limit will be characterized by a linear time trend, and the identification of the key parameter $\beta_1^{'}$ will depend on the difference in the time trend between the control and treatment groups.

Equation (14) also suggests that the identification strategy will be sensitive to model specification. In particular, if the true model is autoregressive, other policies can also generate a difference in the time trend between the control and treatment groups. For instance, suppose the time limit has no effect (i.e., $\beta_0 = \beta_1 = 0$), and consider the following simple data generating process that has one autoregressive lag in y and a treatment group indicator E_i:

$$y_{it} = \lambda_0 + \alpha y_{i,t-1} + \beta_E E_i + \epsilon_{it} \tag{15}$$

It can readily be shown that $E(y_{it}|y_{i0}, E_i) = \alpha^t y_{i0} + \frac{\beta_E}{1-\alpha} E_i - \beta_E E_i \alpha^t$, which implies that the treatment status E_i alone can also result in different time trends between the control and treatment groups via persistent effects on y over time. As a result, the instrumental variable approach requires other information, in particular, multiple variations in the stringency of the time limit (S_{i0}), for more robust identification.

Another challenge of the instrumental variable approach is model misspecification. The simulated stock of remaining months of welfare eligibility Z_{ijt} is assumed to deplete *linearly* over time (equations (12) and (14)). However, as shown in Lemma 1 of the Appendix, the expected stock of remaining months of welfare eligibility depletes in a highly *nonlinear* manner (equation (20)). As a result, the model may produce a biased estimate of the effect of the time limit.

4 Estimation of the Baseline Model

The main challenge to estimating the baseline model in equation (1) is that the lagged dependent variables $y_{i,t-1}, ..., y_{i,t-k}$ and the stock of remaining months of welfare eligibility S_{it} are correlated with unobserved individual effect μ_i. Endogeneity of the above variables are formally taken into account using a generalized method of moments (GMM) estimator. First, rearranging terms in equation (1) and taking first-order difference, the unobserved individual effect μ_i is canceled out:[17]

$$\Delta y_{it} = \sum_{k=1}^{K} \alpha_k \Delta y_{i,t-k} + (\beta_0 - 24\beta_S - 16\beta_A) E_i \Delta\mathbf{1}\{S_{it} < H_{it}\} - \beta_S E_i \mathbf{1}\{S_{i,t-1} < H_{i,t-1}\} y_{i,t-1} + \beta_A E_i \left(\mathbf{1}\{S_{it} < H_{it}\} A_{it} - \mathbf{1}\{S_{i,t-1} < H_{i,t-1}\} A_{i,t-1}\right) + \Delta \boldsymbol{X_{it}} \boldsymbol{\lambda} + \Delta\epsilon_{it}, \tag{16}$$

where $\Delta y_{it} = y_{it} - y_{i,t-1}$, $\Delta\mathbf{1}\{S_{it} < H_{it}\} = \mathbf{1}\{S_{it} < H_{it}\} - \mathbf{1}\{S_{i,t-1} < H_{i,t-1}\}$, $\Delta \boldsymbol{X_{it}} = \boldsymbol{X_{it}} - \boldsymbol{X_{i,t-1}}$, and $\Delta\epsilon_{it} = \epsilon_{it} - \epsilon_{i,t-1}$. Then, in a similar spirit to Anderson and Hsiao (1982) and Arellano and Bond (1991), lags of the dependent variable are used for constructing instruments for the first-difference equation. By the structure of the model, $y_{i0}, y_{i1}, ..., y_{i,t-2}$ are uncorrelated with ϵ_{it} and $\epsilon_{i,t-1}$. Therefore, the population moments of particular interest are:

$$E(y_{i,t-2-m} \Delta\epsilon_{it}) = 0 \qquad \forall m = 0, ..., t-2, \tag{17}$$

$$E(E_i \mathbf{1}\{S_{i0} < H_{i0}\} y_{i,t-2-m} \Delta\epsilon_{it}) = 0 \qquad \forall m = 0, ..., t-2. \tag{18}$$

While the moment conditions in equation (17) are standard in the literature of dynamic panel data models, the set of moments in equation (18) is introduced as a customization for the analysis of time limits. It is important for the identification of the stock coefficient β_S because it captures the difference in behavioral dynamics between the control and treatment groups. In particular, in the treatment group, participating in welfare now has the additional effect of bringing the individual one period closer to exhausting her welfare eligibility. The moment conditions reflect this difference in dynamics by interacting the treatment group indicator with the lagged dependent variable.[18]

The corresponding sample moments for individual i are constructed as $\sum_{t=1+m}^{T} y_{i,t-1-m} \Delta\epsilon_{it}$, where $m = 1, ..., \bar{m}$ (e.g., Anderson and Hsiao (1982)). These moments serve to identify the state dependence parameters α_k. Similarly, sample moments $E_i \mathbf{1}\{S_{i0} < H_{i0}\} \sum_{t=1+m}^{T} y_{i,t-1-m} \Delta\epsilon_{it}$, where $m = 1, ..., \bar{m}$, serve to identify the stock effect β_S. In the baseline model, $\bar{m}$ is set to be

[17] It can readily be shown that $\mathbf{1}\{S_{it} < H_{it}\} S_{it} - \mathbf{1}\{S_{i,t-1} < H_{i,t-1}\} S_{i,t-1} = -y_{i,t-1} \mathbf{1}\{S_{i,t-1} < H_{i,t-1}\}$, which leads to the term associated with β_S in equation (16).

[18] In equation (18) and estimation of the model, $\mathbf{1}\{S_{i0} < H_{i0}\}$ replaces $\mathbf{1}\{S_{it} < H_{it}\}$ as an instrument. For more details, see the instrument matrix in equations (23) and (24) of the Appendix. In the data, there are only 15 observations where $\Delta\mathbf{1}\{S_{it} < H_{it}\} \neq 0$, and $\mathbf{1}\{S_{it} < H_{it}\}$ is equal to $\mathbf{1}\{S_{i0} < H_{i0}\}$ for most individuals.

12. Therefore, there are $12 \times 2 = 24$ instruments described above.[19] In addition, the length of the autoregressive lag K is set to be 3; both the choice of the number of instruments and the autoregressive lag will be discussed in detail in the Results section. Sample moments for other parameters in the model, the full instrument matrix, and the formula of the GMM estimator are defined formally in the Appendix.

5 Results from the Baseline Sample

Comparison of Model Estimates. Table II reports estimates of the baseline model in equation (1), and estimates from GM and Mazzolari models. Column 1 presents the GM model (equation (3)) that is estimated on data from the first six months after random assignment only. The proximity of the sample period to the time of random assignment implies that the GM effect estimate should be relatively unbiased. The GM coefficient (β_A) is 0.003, but is statistically insignificant at the 10 percent level. This finding is consistent with the sensitivity analysis in Grogger and Michalopoulos (2003).[20] The coefficient for treatment status (β_E) is 0.062, which implies that other features of FTP (i.e., improved financial incentives and enhanced services) increase the welfare participation rate by 6.2 percentage points.

Column 2 presents the GM model that is estimated on the full baseline sample. This specification follows the baseline model in Grogger and Michalopoulos (2003). The GM coefficient β_A is 0.014, and is statistically significant at the 1 percent level.[21] The estimate suggests that if the age of the youngest child is smaller by one year, the time limit will reduce the welfare participation rate by an additional 1.4 percentage points. According to equation (10), this is likely an overestimate of the true GM effect. The treatment group coefficient β_E is 0.103, which is 4 percentage points larger than the estimate in Column 1. However, since the treatment group coefficient partly subsumes the stock effect of the time limit, it cannot be interpreted as the effects of other features of FTP

[19] There is no universal rule for the choice of instruments, which is subject to many empirical considerations including sample size. For instance, see Ziliak (1997) who compares the empirical performance of GMM and other instrumental variable estimators for panel data in an empirical application to life-cycle labor supply.

[20] In the sensitivity analysis, their model uses only the first 6 months following random assignment for estimation. They also find that the GM coefficient is statistically insignificant at the 10 percent level.

[21] The GM coefficient is larger than in Grogger and Michalopoulos (2003), which is likely because their analysis sample also contains individuals who are randomly assigned in later periods. We also estimate a model that follows their specification, using welfare use and employment during the pre-random-assignment period as additional regressors. Although the GM coefficient becomes smaller, it is still larger than their estimate.

per se. Column 3 extends the GM model by including measures of the stock effect via coefficients β_0 and β_S. The model is estimated using OLS instead of GMM. The stock coefficient β_S is has the wrong sign at -0.044. This is because the stock of remaining months of welfare eligibility S_{it} is *negatively* correlated with the unobserved effect μ_i, which creates a downward bias in the estimate of the stock coefficient.

Column 4 presents estimates from the Mazzolari model (equation 11). The coefficients β_0 and β_1, which capture the effect of the time limit, are -0.184 and 0.424, respectively, and have expected signs. The estimate for β_0 implies that the time limit reduces the welfare participation rate by 18.4 percentage points for an individual who almost exhausts her stock of welfare eligibility under the time limit. The estimate for β_1 implies that the time limit has a weaker effect if the stock S_{it} is larger. The model also predicts that the time limit will increase welfare participation if the stock S_{it} is larger than 43.4 percent of the length of the remaining time horizon H_{it} (i.e., when $S_{it}/H_{it} > 0.184/0.424 = 0.434$).

Column 5 extends the Mazzolari model by including lagged welfare participation as regressors. [22] The estimates for β_0 and β_1, as well as the treatment group coefficient β_E, all become close to zero and are statistically insignificant at the 10 percent level. Since the instruments of the model exclude individual-level information that is important for identification, it becomes more difficult to disentangle the policy effects under a more general model specification.

Column 6 presents estimates from the baseline dynamic model that is estimated by GMM (equation (1)). The estimates suggest that the stock and GM effects of the time limit, as well as other aspects of FTP, all play a role in affecting welfare participation. The estimate for β_0 implies that the effect of the time limit is insignificant for a treatment group individual whose stock S_{it} is 24 months and whose youngest child is just under 16 years of age. The estimate for the stock coefficient β_S is 0.006, and is statistically significant at the 10 percent level. This implies that the time limit tends to reduce welfare participation by an extra 0.6 percentage points, if the individual has one fewer remaining month of welfare eligibility. In particular, all else being equal, for an individual who has only one month of stock left, the time limit will reduce her welfare participation by $0.6 \times 23 = 13.8$ percentage points. The estimate for the GM coefficient β_A is

[22] The extended model uses the same instruments as the original specification to account for the endogeneity of the stock S_{it}. It uses moment conditions in equation (17) to construct instruments for lagged welfare participation. See the Appendix for further details regarding model estimation.

0.003, which is similar in size to the estimate in Column 1, but is statistically significant at the 10 percent level. The estimate suggests that the time limit reduces welfare participation by an extra 0.3 percentage points, if the individual's youngest child is one *year* younger. The treatment group coefficient β_E is 0.034, and is statistically significant at the 10 percent level. This suggests that other aspects of FTP *immediately* increases welfare participation by 3.4 percentage points. The first- and second-order state dependence coefficients (α_1 and α_2) are 0.660 and 0.116, respectively, and both are statistically significant at the 1 percent level. By contrast, higher order dynamics seem to be inconsequential, as the third-order state dependence coefficient (α_3) is only 0.006, and is statistically insignificant at the 10 percent level.

The last rows of Column 6 report results from two common model specification tests for dynamic panel data models. The first test is the Sargan test of overidentifying restrictions, which tests for the validity of instruments in the GMM estimator. Under the null hypothesis (i.e., instruments are valid), the test statistic follows a chi-square distribution with a degree of freedom that is equal to the number of overidentifying restrictions. In the baseline model, the Sargan statistic is 23.36 with 20 degrees of freedom, and the null hypothesis is not rejected at the 10 percent level. The second test is the Arellano-Bond test of serial correlation in the transitory errors (ϵ_{it}) of the model (Arellano and Bond (1991)).[23] Under the null hypothesis (i.e., no serial correlation in ϵ_{it}), the test statistic will follow a standard normal distribution asymptotically; in the alternative hypothesis, the transitory errors are serially correlated, which will affect the validity of the instruments. The test statistic is 0.116, and the null hypothesis is not rejected at the 10 percent level, which is consistent with results from the Sargan test.

Comparison of Policy Effects. Does the dynamic panel data model generate a different effect of the time limit? Table III reports the simulated policy effects during the sample period based on the model estimates. The effects are computed by comparing the simulated outcomes between the baseline policy scenario in the *treatment group*, and the counterfactual scenario in which a policy is "shut off" by setting the relevant coefficient estimate(s) to zero. The first row reports simulation results from the baseline model.[24] The time limit reduces welfare participation by an average of

[23] Arellano and Bond (1991) show that both specification tests are asymptotically equivalent.

[24] The outcomes are simulated during the sample period conditional on unobserved effects, which are computed from GMM residuals.

14.6 percentage points during the sample period. Around 60 percent (or 9.0 percentage points) of this reduction are due to the GM effect via coefficient β_A, and 40 percent (or 5.6 percentage points) are due to the stock effect via coefficients β_0 and β_S. Other aspects of FTP are predicted to increase welfare participation by an average of 10.7 percentage points via coefficient β_E.[25]

The second row of the table reports simulation results from the GM model (Column 2 in Table II). The time limit is predicted to reduce welfare participation by an average of 13.8 percentage points during the sample period. This is all due to the GM effect by construction of the model. Since most of the stock effect are subsumed into the GM coefficient, the model overstates the GM effect and slightly understates the overall effect of the time limit. Other aspects of FTP are predicted to increase welfare participation by 10.3 percentage points.

The third row of the table reports simulation results from the Mazzolari model (Column 4 in Table II). The time limit is predicted to reduce welfare participation by an average of 11.4 percentage points during the sample period. Interestingly, most of this reduction are due to the GM effect, and the stock effect constitutes 0.7 percentage points of the reduction only.[26] This result is qualitatively similar to Mazzolari (2007), who also finds a small stock effect. Other aspects of FTP are predicted to increase welfare participation by 9.1 percentage points.

The fourth row of the table reports simulation results from an extended version of the GM model. In this model, three orders of lagged welfare participation are included as additional regressors (i.e., $y_{i,t-1}$, $y_{i,t-2}$, $y_{i,t-3}$).[27] The model accurately predicts the magnitude of the GM effect (9.4 percentage points); however, since the stock effect is zero by construction, the model severely understates the overall effect of the time limit. As a result, the effect of other aspects of FTP is also understated (8.1 percentage points).

The bottom panel of the table reports the simulated policy effects at months 5, 11, and 23 after random assignment, respectively. The baseline model predicts that the stock effect is initially negligible, but it becomes substantial by the end of the sample period. For instance, at month 5,

[25]Since the model is dynamic, the effect of implementing two policies simultaneously can be different from the sum of the effects of implementing each policy separately. The counterfactual on other aspects of FTP does not entail the interaction coefficient between treatment status and the youngest child under age 3 indicator.

[26]In this model, the stock effect is obtained by comparing the predicted outcomes in the treatment group between the baseline scenario and when the stock S_{it} is restricted to 24. The GM effect is then computed as the difference between the overall and the stock effect of the time limit.

[27]The extended model uses moment conditions in equation (17) to construct instruments for lagged welfare participation. The estimates are available upon request. See the Appendix for further details regarding model estimation.

the stock effect reduces welfare participation by 0.1 percentage points only. As time progresses, individuals became more subject to the stock effect as they deplete their stock of remaining months of welfare eligibility. At month 23, the stock effect reduces welfare participation by 12.5 percentage points, while the GM effect in the same month is only 10.5 percentage points. The extended GM model predicts the size of the GM effect accurately over time; however, since the stock effect is zero by construction, the model tends to understate the overall effect of the time limit as time progresses.

Additional Specifications. Table IV provides additional evidence for the effects of FTP under a variety of model specifications. In Column 1, the baseline model is expanded to include work experience (ω_{it}), as well as contemporaneous and three lagged indicators of employment status (x_{it}, $x_{i,t-1}$, $x_{i,t-2}$, $x_{i,t-3}$) as additional regressors. The main objective of this exercise is to control for work experience. As treatment group individuals may accumulate a higher level of work experience than control group individuals, this may potentially affect the model estimates in a complicated way. Since the work-related variables are endogenous, additional instruments are constructed in a similar spirit to moments conditions in equation (17) using individual-level employment data. The model and the corresponding instrument matrix for the GMM estimator are defined formally in the Appendix. After controlling for work-related variables, both the coefficient for the stock effect β_S and the GM coefficient β_A remain similar in size, and the treatment group coefficient β_E becomes slightly smaller at 0.022. The estimates also reveal interesting relationships between welfare participation and labor market outcomes. For instance, all else being equal, welfare participation becomes lower by 0.7 percentage points when an individual accumulates an extra month of work experience. In addition, all else being equal, an individual is 13.1 percentage points less likely to participate in welfare if she works. Both results are consistent with the structure of the welfare program, which provides partial or no payments to workers who have earnings beyond the level designated by the earnings disregard.

Column 2 extends the baseline model by allowing for the stock effect to be a piecewise linear function of the remaining stock S_{it} (e.g., Ribar et al. (2008)). Under this specification, if the individual has more than six remaining months of welfare eligibility, the marginal effect of the stock on welfare participation is β_S. Otherwise, the marginal effect is $\beta_S + \beta_{S1}$, where β_{S1} captures the piecewise linear effect. An extra instrument is required to estimate β_{S1}; the model and its corresponding instrument matrix are defined formally in the Appendix. The results are qualitatively similar to the baseline model. The estimate for β_S is 0.001 with a standard error of 0.004, and the estimate for β_{S1} is 0.033 with a standard error of 0.018. The GM coefficient and treatment group coefficient are 0.003 and 0.026, respectively, which are similar to the estimates from the baseline model.

In Column 3, the model is assumed to exhibit first-order state dependence only, and higher-order lags of the dependent variable ($y_{i,t-2}$ and $y_{i,t-3}$) are excluded from the set of regressors. Both the stock coefficient β_S and the GM coefficient β_A remain similar in size to estimates from the baseline model, while the treatment group coefficient becomes slightly larger at 0.041. The coefficient for the first-order autoregressive lag (α_1) also becomes larger at 0.731, as it picks up the effects from the omitted second-order lag, which was statistically significant in the baseline model. Both the Sargan test and the Arellano-Bond test reject the null hypothesis at the 1 percent level, implying that the model is misspecified. This result, which is not surprising, illustrates that serial correlation of transitory errors is closely related to the omission of significant lagged dependent variables in the model.

The last two columns of the table examine whether the estimation results are sensitive to the instruments used. In the baseline model, there are 12 lagged dependent variables involved in instrument construction (i.e., $y_{i,t-2}$, ..., $y_{i,t-1-\bar{m}}$ where $\bar{m} = 12$). Column 4 uses six lagged dependent variables for instrument construction (i.e., $\bar{m} = 6$), and Column 5 uses 18 lagged dependent variables for instrument construction (i.e., $\bar{m} = 18$). In both cases, the estimation results are very similar to the baseline model.

6 Results from the Disadvantaged Sample

As discussed previously, individuals who are considered as disadvantaged are far more likely to participate in welfare than individuals in the baseline sample. These individuals may respond to FTP policies differently. To examine whether the effects of FTP are different among these individuals, the models in previous sections are re-estimated using data from the disadvantaged sample.[28] Estimation results from selected models are reported in Table V.

Column 1 of the table reports estimates from the GM model. The GM coefficient β_A is 0.012, which is slightly smaller than the estimate from the baseline sample (Column 2 in Table II). By contrast, the treatment group coefficient is 0.158, which is almost 50 percent larger than the estimate from the baseline sample.

Column 2 reports estimates from the baseline model. As expected, the estimate for β_0 implies that the time limit has an insignificant effect on a treatment group individual whose remaining stock S_{it} is 36 months and whose youngest child is just under 15 years of age. The stock coefficient β_S is 0.003, which is substantially smaller in size than the estimate from the baseline sample (Column 6 in Table II), but is statistically significant at the 5 percent level. Therefore, disadvantaged individuals are less responsive to the stock of remaining welfare eligibility under the time limit. By contrast, the GM coefficient is 0.004, which is similar to the estimate from the baseline sample. The treatment group coefficient is 0.08, which is more than twice as large as the estimate from the baseline sample. This suggests that other aspects of FTP have a disproportionately large positive effect on welfare participation among disadvantaged individuals.

Column 3 presents estimates from an extended baseline model that controls for work-related variables. The key estimates including the stock coefficient β_S, the GM coefficient β_A, and the treatment group coefficient β_E are all similar in size to estimates in Column 2. Interestingly, unlike the estimates from the baseline sample (Column 1 in Table IV), neither work experience nor employment status have a significant effect on welfare participation. This suggests that the relationship between welfare participation and labor market outcomes is very weak among disadvantaged individuals. In particular, attachment to the labor market does not induce them to leave welfare. Since disadvantaged individuals typically receive low earnings when they work, this result is consistent with the structure of the earnings disregard in the welfare program, which keeps welfare payments in full to workers who have low earnings.

Comparison of Policy Effects. Using the same approach as the baseline sample analysis, we simulate the policy effects during the sample period based on several models that are re-estimated using data from the disadvantaged sample. The results are reported in Table VI.

[28] As noted previously, the model specifications are the same as the model for the baseline sample after adjustments in the terms involving the stock (replace 24 months by 36) and the age of the youngest child (replace 16 years by 15).

The first row of the table reports simulation results from the baseline model. The model predicts that the time limit reduces welfare participation by an average of 19.9 percentage points during the sample period. The GM and stock effects of the time limit constitute 10.9 and 9.0 percentage points of the reduction, respectively. Other aspects of FTP are predicted to increase welfare participation by 23 percentage points. The above policy effects are much larger in size than estimates from the baseline sample. Interestingly, despite the smaller stock coefficient β_S, which reflects a lower *marginal* behavioral response to the stock of remaining welfare eligibility, the size of the *total* stock effect among disadvantaged individuals is substantially larger than individuals in the baseline sample. This is because disadvantaged individuals have a much higher welfare participation rate in general, which makes them deplete their stock of remaining welfare eligibility at a much faster rate than individuals in the baseline sample. As a result, these individuals are more subject to the stock effect.

The second row of the table reports simulation results from the GM model. The model predicts that the time limit reduces welfare participation by only 11.1 percentage points during the sample period. Therefore, in the disadvantaged sample, the GM model understates the overall effect of the time limit by approximately 40 percent. Simulation results from the next two rows, which come from the Mazzolari model and the extended GM model (with lagged welfare participation), respectively, show a similar picture.[29]

The bottom panel of the table reports the simulated policy effects at months 5, 11, and 23 after random assignment, respectively. In the baseline model, the stock effect of the time limit already reduces welfare participation at month 5 by 3.0 percentage points; at month 23, it is as large as 15.8 percentage points. By contrast, although the extended GM model incorporates lagged dependent variables as regressors, its prediction of the effect of the time limit is much less dynamical, and it tends to understate the overall effect of the policy as time progresses.

7 Conclusions

In this paper, we analyzed the effects of welfare time limits using a dynamic panel data model and data from a randomized policy experiment (FTP). The model incorporated dynamic welfare par-

[29]The estimates from both models are available upon request.

ticipation behavior under time limits, and formally took into account of unobserved heterogeneity using panel data. It was shown analytically and empirically that a dynamic model specification had major advantages for the identification and correct measurement of the effects of time limits. Estimation was performed using a generalized method of moments (GMM) estimator, with instruments constructed from the panel features of the data in a way that was consistent with the literature of dynamic panel data models.

We found that around 40 percent of the anticipatory effect of the time limit were due to individuals reducing their welfare use when their stock of remaining welfare eligibility became lower. Due to a sizable stock effect, the effect of the time limit was highly dynamical and exhibited a strong increase over time. Somewhat surprisingly, we also found that the time limit led to a larger reduction in welfare use among disadvantaged individuals. Although their *marginal* behavioral response to the stock of remaining welfare eligibility was lower, their stocks were depleted at a much faster rate due to a high dependence on welfare, which resulted in a larger *total* stock effect. On the other hand, other aspects of FTP, which included financial work incentives and enhanced employment services, had a strong positive effect on welfare participation among disadvantaged individuals. A potential factor that might have played a role was the relationship between welfare participation and labor market outcomes, which was found to be much weaker among disadvantaged individuals than individuals in the baseline sample.

The results in this paper have interesting implications on the dynamics of welfare caseloads especially during the post-TANF period. Relatively few studies have investigated the general policy implications of dynamic welfare participation behavior (e.g., Ziliak et al. (2000), Keane and Wolpin (2002), Haider and Klerman (2005)). As Klerman and Haider (2004) show using California data, which had no time limit, a relatively simple individual-level model of welfare participation behavior can generate rich dynamics in aggregate welfare caseloads. As states began to implement time limits in the late 1990s, welfare caseloads also experienced a steep decline during the same period. As a potentially fruitful direction of research, it would be interesting to examine whether the caseload data during the post-TANF period was subject to different dynamics, and whether the difference would be consistent with a behavioral model of time limits that was presented in this paper.

8 Appendix 1: Technical Results and Proofs

The results in this section provide a theoretical basis for the comparison between the baseline model and existing regression approaches in Section 3. The analysis starts from the stylized version of the baseline model in equation (5).

Lemma 1. *Suppose* $0 \leq \beta_S < 1$. *Then, the conditional expectations of* y_{it} *and* S_{it} *for treatment group individuals at* $t = 1, 2, ..., T$ *are*

$$E(y_{it}|A_i, E_i = 1) = (\lambda_0^{'} + \beta_S S_{i0} + (\beta_A + \lambda_A)A_i)(1 - \beta_S)^t, \tag{19}$$

$$E(S_{it}|A_i, E_i = 1) = S_{i0} - (\lambda_0^{'} + \beta_S S_{i0} + (\beta_A + \lambda_A)A_i)\sum_{k=0}^{t-1}(1 - \beta_S)^k. \tag{20}$$

Proof *Equation (5) can be rewritten as*

$$y_{it} = \tilde{\lambda}_0^{'} + \tilde{\beta}_S E_i \tilde{S}_{it} + (\beta_A E_i + \lambda_A)A_i + \epsilon_{it},$$

$$\tilde{S}_{it} = \sum_{k=0}^{t-1} y_{ik}, \quad \tilde{S}_{i0} = 0,$$

where $\tilde{\lambda}_0^{'} = \lambda_0^{'} + \beta_S E_i S_{i0}$ *and* $\tilde{\beta}_S = -\beta_S$. *We need to show by induction that for* $t = 1, 2, ..., T$, $E(y_{it}|A_i, E_i = 1) = (\tilde{\lambda}_0^{'} + (\beta_A + \lambda_A)A_i)(1 + \tilde{\beta}_S)^t$ *and* $E(\tilde{S}_{it}|A_i, E_i = 1) = (\tilde{\lambda}_0^{'} + (\beta_A + \lambda_A)A_i))\sum_{k=0}^{t-1}(1 + \tilde{\beta}_S)^k$. *It is straightforward to show that the statements are true for* $t = 1$. *Suppose it is true that* $E(y_{ir}|A_i, E_i = 1) = (\tilde{\lambda}_0^{'} + (\beta_A + \lambda_A)A_i))(1 + \tilde{\beta}_S)^r$ *and* $E(\tilde{S}_{ir}|A_i, E_i = 1) = (\tilde{\lambda}_0^{'} + (\beta_A + \lambda_A)A_i))\sum_{k=0}^{r-1}(1 + \tilde{\beta}_S)^k$ *for some* $r \geq 1$. *Then for* $t = r + 1$, *we have* $E(\tilde{S}_{i,r+1}|A_i, E_i = 1) = E(y_{ir} + \tilde{S}_{ir}|A_i, E_i = 1) = (\tilde{\lambda}_0^{'} + (\beta_A + \lambda_A)A_i)\sum_{k=0}^{r}(1 + \tilde{\beta}_S)^k$, *and*

$$\begin{aligned} E(y_{i,r+1}|A_i, E_i = 1) &= \tilde{\lambda}_0^{'} + \tilde{\beta}_S(\tilde{\lambda}_0^{'} + (\beta_A + \lambda_A)A_i)\sum_{k=0}^{r}(1 + \tilde{\beta}_S)^k + (\beta_A + \lambda_A)A_i \\ &= (\tilde{\lambda}_0^{'} + (\beta_A + \lambda_A)A_i)(1 + \tilde{\beta}_S \sum_{k=0}^{r}(1 + \tilde{\beta}_S)^k) \\ &= (\tilde{\lambda}_0^{'} + (\beta_A + \lambda_A)A_i)(1 + \tilde{\beta}_S)^{r+1}. \end{aligned}$$

Therefore the statements are true for $t = 1, ..., T$. *QED.*

The following result follows directly from equation (19).

Lemma 2. *Consider two different ages of the youngest child in the family, a and $a^{'}$. Then,*

$$E(y_{it}|A_i = a^{'}, E_i = 1) - E(y_{it}|A_i = a, E_i = 1) = (\beta_A + \lambda_A)(a^{'} - a)(1 - \beta_S)^t.$$

The following proposition shows that the GM effect obtained from the GM model ($\tilde{\beta}_A$) is generally different from the true GM effect (β_A).

Proposition 1. $\tilde{\beta}_A - \beta_A = (a^{'} - a)^{-1}(\beta_A + \lambda_A)\left((1 - \beta_S)^t - 1\right)$.

Proof. *From the definitions of $\tilde{\beta}_A$ and β_A in equations (8) and (9), we have*

$$\begin{aligned} \tilde{\beta}_A - \beta_A &= (a^{'} - a)^{-1}\Big[\Big(E(y_{it}|A_i = a^{'}, E_i = 1) - E(y_{it}|A_i = a, E_i = 1)\Big) - \\ &\qquad \Big(E(y_{it}|A_i = a^{'}, E_i = 1, S_{it} = s) - E(y_{it}|A_i = a, E_i = 1, S_{it} = s)\Big)\Big] \\ &= (a^{'} - a)^{-1}(\beta_A + \lambda_A)(1 - \beta_S)^t - (a^{'} - a)^{-1}(\beta_A + \lambda_A) \\ &= (a^{'} - a)^{-1}(\beta_A + \lambda_A)\left((1 - \beta_S)^t - 1\right), \end{aligned} \tag{21}$$

where the second equality comes from Lemma 2 and equation (7). QED.

The following proposition provides another representation for $\tilde{\beta}_A - \beta_A$.

Proposition 2. $\tilde{\beta}_A - \beta_A = (a^{'} - a)^{-1}\beta_S\Big(E(S_{it}|A_i = a^{'}, E_i = 1) - E(S_{it}|A_i = a, E_i = 1)\Big)$.

Proof. *We first express the conditional expectation of y_{it} as a function of the conditional expectation of S_{it}:*

$$\begin{aligned} E(y_{it}|A_i = a, E_i = 1) &= \int E(y_{it}|A_i = a, E_i = 1, S_{it} = s) f(s|A_i = a, E_i = 1) ds \\ &= \int (\lambda_0^{'} + (\beta_A + \lambda_A)a + \beta_S s) f(s|A_i = a, E_i = 1) ds \\ &= \lambda_0^{'} + (\beta_A + \lambda_A)a + \beta_S \int s f(s|A_i = a, E_i = 1) ds \\ &= \lambda_0^{'} + (\beta_A + \lambda_A)a + \beta_S E(S_{it}|A_i = a, E_i = 1), \end{aligned}$$

where $f(.)$ denotes the probability density function of stock s. Then, using the above expression for $A_i = a^{'}$ and $A_i = a$, we have

$$\begin{aligned} &E(y_{it}|A_i = a^{'}, E_i = 1) - E(y_{it}|A_i = a, E_i = 1) \\ &= (\beta_A + \lambda_A)(a^{'} - a) + \beta_S\Big(E(S_{it}|A_i = a^{'}, E_i = 1) - E(S_{it}|A_i = a, E_i = 1)\Big). \end{aligned}$$

The proposition then follows by substituting the above expression and equation (7) into equation (21). QED.

9 Appendix 2: GMM Estimator

GMM Estimator for the Baseline Model. For individual i, the instrument submatrix related to the moments conditions in equations (17) and (18) is $[Z_{iy} \quad E_i\mathbf{1}\{S_{i0} < H_{i0}\}Z_{iy}]$, where

$$Z_{iy} = \begin{bmatrix} y_{i2} & y_{i1} & y_{i0} & 0 & \cdots & 0 \\ y_{i3} & y_{i2} & y_{i1} & y_{i0} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ y_{i,T-2} & y_{i,T-3} & y_{i,T-4} & y_{i,T-5} & \cdots & y_{i,T-1-\bar{m}} \end{bmatrix}_{(T-3)\times\bar{m}}, \tag{22}$$

and $E_i\mathbf{1}\{S_{i0} < H_{i0}\}Z_{iy}$ is a $(T-3)\times\bar{m}$ matrix that is equal to the product of scalar $E_i\mathbf{1}\{S_{i0} < H_{i0}\}$ and matrix Z_{iy}. The instrument submatrix related to other regressors in the model is

$$Z_{ix} = \begin{bmatrix} E_i\mathbf{1}\{S_{i0} < H_{i0}\} & E_i\mathbf{1}\{S_{i0} < H_{i0}\}A_{i0} & E_i & \boldsymbol{X}_{i0} \\ E_i\mathbf{1}\{S_{i0} < H_{i0}\} & E_i\mathbf{1}\{S_{i0} < H_{i0}\}A_{i1} & E_i & \boldsymbol{X}_{i1} \\ \vdots & \vdots & \vdots & \vdots \\ E_i\mathbf{1}\{S_{i0} < H_{i0}\} & E_i\mathbf{1}\{S_{i0} < H_{i0}\}A_{i,T-1} & E_i & \boldsymbol{X}_{i,T-1} \end{bmatrix}_{T\times(K_x+3)}. \tag{23}$$

The full instrument matrix for individual i is

$$Z_i = \left[\begin{array}{cc|c} Z_{iy} & E_i\mathbf{1}\{S_{i0} < H_{i0}\}Z_{iy} & 0 \\ \hline 0 & 0 & Z_{ix} \end{array}\right], \tag{24}$$

where the number of columns in Z_i is $2\bar{m}+K_x+3$, which represents the total number of instruments in the model.[30] The model is overidentified, and the number of overidentifying restrictions equals $2\bar{m} + K_x + 3 - (K + 1 + K_x + 3) = 2\bar{m} - K - 1$.

The GMM estimator is described as follows. For individual i at month t, denote the residual by $e_{it} = y_{it} - \tilde{\boldsymbol{X}}_{it}\boldsymbol{b}$, where $\tilde{\boldsymbol{X}}_{it}$ is the vector of all regressors on the right hand side of equation (1) and $\boldsymbol{b}$ is the vector of coefficient estimates. Denote the residual vector of individual i by $\tilde{\boldsymbol{e}}_i = [\Delta e_{i2}...\Delta e_{iT}|e_{i1}...e_{iT}]'$, where $\Delta e_{it} = e_{it} - e_{i,t-1}$. The GMM estimator is the solution to the

[30] The columns in the instrument submatrix Z_{ix} have effectively the same function as regressors in the specifications of Grogger and Michalopoulos (2003) and Mazzolari (2007). In the GMM literature, they are often referred to as "level" instruments (e.g., Arellano and Bover (1995)).

following objective function:

$$\min_{\boldsymbol{b}} \left(\frac{1}{N} \sum_{i=1}^{N} Z_i' \tilde{\boldsymbol{e}}_i \right)' A \left(\frac{1}{N} \sum_{i=1}^{N} Z_i' \tilde{\boldsymbol{e}}_i \right), \tag{25}$$

where A is the weighting matrix.

Let $y_i = [\Delta y_{i2} ... \Delta y_{iT} | y_{i1} ... y_{iT}]'$, $\tilde{X}_i = [\Delta \tilde{\boldsymbol{X}}_{\boldsymbol{i2}} ... \Delta \tilde{\boldsymbol{X}}_{\boldsymbol{iT}} | \tilde{\boldsymbol{X}}_{\boldsymbol{i1}} ... \tilde{\boldsymbol{X}}_{\boldsymbol{iT}}]'$, $D_{zy} = \frac{1}{N} \sum_{i=1}^{N} Z_i' \tilde{y}_i$, and $D_{zx} = \frac{1}{N} \sum_{i=1}^{N} Z_i' \tilde{X}_i$. The GMM estimator has a closed form solution which is

$$b_{GMM} = \left(D_{zx}' A D_{zx} \right)^{-1} D_{zx}' A D_{zy}. \tag{26}$$

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 the entry in row i and column j as $h_{ij} = 2$ if $i = j$, $h_{ij} = 1$ if $|i - j| = 1$, $h_{ij} = 0$ if $|i - j| > 1$. The feasible variance of the one-step estimator is

$$Var(b_{GMM1}) = \frac{1}{N} (D_{zx}' A D_{zx})^{-1} D_{zx}' A \hat{V} A D_{zx} (D_{zx}' A D_{zx})^{-1}, \tag{27}$$

where $\hat{V} = \frac{1}{N} \sum_{i=1}^{N} Z_i' \hat{\tilde{e}}_i \hat{\tilde{e}}_i' Z_i$, and $\hat{\tilde{e}}_i$ equals the estimated GMM residual vector.

We use the two-step GMM estimator, which has the optimal feasible weighting matrix $A = \hat{V}^{-1}$. The feasible variance of the two-step estimator is

$$Var(b_{GMM2}) = \frac{1}{N} (D_{zx}' \hat{V}^{-1} D_{zx})^{-1}. \tag{28}$$

Controlling for Work Experience. We discuss an extension of the baseline model that controls for work experience and other work-related variables (e.g., Column 1 in Table IV). For the baseline sample, the model is defined formally as

$$y_{it} = \lambda_0 + \sum_{k=1}^{K} \alpha_k y_{i,t-k} + \sum_{k=0}^{K} \gamma_k x_{i,t-k} + \gamma_w \omega_{it} + \beta_0 E_i \mathbf{1}\{S_{it} < H_{it}\} + \beta_S E_i \mathbf{1}\{S_{it} < H_{it}\}(S_{it} - 24) + \beta_A E_i \mathbf{1}\{S_{it} < H_{it}\}(A_{it} - 16) + \beta_E E_i + \boldsymbol{X_{it}} \boldsymbol{\lambda} + \mu_i + \epsilon_{it}, \tag{29}$$

where x_{it} is a employment status indicator for individual i at month t (=1 if she works, =0 otherwise), and $\omega_{it} = \sum_{k=0}^{t-1} x_{ik}$ is her cumulative months of work since random assignment.[31] In general, we should expect that there is a negative relationship between welfare participation and work-related variables, that is, the expected signs of γ_k and γ_w are negative. Similar to equation (16), after taking first-order difference the above model becomes

$$\begin{aligned}\Delta y_{it} = &\sum_{k=1}^{K} \alpha_k \Delta y_{i,t-k} + \sum_{k=0}^{K} \gamma_k \Delta x_{i,t-k} + \gamma_w x_{i,t-1} + (\beta_0 - 24\beta_S - 16\beta_A) E_i \mathbf{\Delta 1}\{S_{it} < H_{it}\} - \\ &\beta_S E_i \mathbf{1}\{S_{i,t-1} < H_{i,t-1}\} y_{i,t-1} + \beta_A E_i \left(\mathbf{1}\{S_{it} < H_{it}\} A_{it} - \mathbf{1}\{S_{i,t-1} < H_{i,t-1}\} A_{i,t-1} \right) + \\ &\Delta \boldsymbol{X_{it}} \boldsymbol{\lambda} + \Delta \epsilon_{it},\end{aligned} \tag{30}$$

where $\Delta x_{it} = x_{it} - x_{i,t-1}$. To address the endogeneity of work-related variables $x_{i,t-k}$ and ω_{it}, the following population moments are used:

$$E(x_{i,t-2-m} \Delta \epsilon_{it}) = 0 \qquad m = 0, ..., t-2. \tag{31}$$

Then, similar to equation (22), the corresponding instrument submatrix is defined as

$$Z_{iw} = \begin{bmatrix} x_{i2} & x_{i1} & x_{i0} & 0 & ... & 0 \\ x_{i3} & x_{i2} & x_{i1} & x_{i0} & ... & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ x_{i,T-2} & x_{i,T-3} & x_{i,T-4} & x_{i,T-5} & ... & x_{i,T-1-\bar{m}} \end{bmatrix}_{(T-3)\times \bar{m}} . \tag{32}$$

The full instrument matrix for individual i is

$$Z_i = \left[\begin{array}{ccc|c} Z_{iy} & E_i \mathbf{1}\{S_{i0} < H_{i0}\} Z_{iy} & Z_{iw} & 0 \\ \hline 0 & 0 & 0 & Z_{ix} \end{array} \right], \tag{33}$$

where the number of columns in Z_i is $3\bar{m} + K_x + 3$, and the number of overidentifying restrictions is $3\bar{m} + K_x + 3 - [(K+1) + (K+2) + (K_x+3)] = 3\bar{m} - 2K - 3$.

[31] The data on work-related variables comes from (calendar) quarterly administrative records on employment in the FTP public use file. The data is converted to a monthly level for analysis.

Piecewise Linear Effects. We discuss an extension of the baseline model that allows for piecewise linear stock effects (e.g., Column 2 in Table IV). For the baseline sample, the model is defined formally as

$$
\begin{aligned}
y_{it} = \lambda_0 + \sum_{k=1}^{K} \alpha_k y_{i,t-k} + \beta_0 E_i \mathbf{1}\{S_{it} < H_{it}\} + \beta_S E_i \mathbf{1}\{S_{it} < H_{it}\}(S_{it} - 24) + \\
\beta_{S1} E_i \mathbf{1}\{S_{it} < H_{it}\} \min\{S_{it} - \bar{s}, 0\} + \beta_A E_i \mathbf{1}\{S_{it} < H_{it}\}(A_{it} - 16) + \beta_E E_i + \boldsymbol{X_{it}\lambda} + \mu_i + \epsilon_{it}.
\end{aligned}
\tag{34}
$$

If S_{it} is larger than $\bar{s}$, the marginal effect of S_{it} will be β_S, otherwise the marginal effect will be $\beta_S + \beta_{S1}$. The model is estimated using an expanded instrument matrix:

$$
Z_i = \left[\begin{array}{ccc|c} Z_{iy} & E_i \mathbf{1}\{S_{i0} < H_{i0}\} Z_{iy} & E_i \mathbf{1}\{S_{i0} < H_{i0}\} \sum_{j=1}^{\bar{m}} Z_{iys}[*j] & 0 \\ \hline 0 & 0 & 0 & Z_{ix} \end{array}\right]. \tag{35}
$$

The extra elements are described as follows. Let matrix Z_{iys} be the entrywise product of Z_{iy} and Z_{is} (i.e., $Z_{iys} \equiv Z_{iy} \circ Z_{is}$), where Z_{is} is a matrix with the following indicator functions as entries:

$$
Z_{is} = \begin{bmatrix} \mathbf{1}(S_{i2} < \bar{s}) & \mathbf{1}(S_{i1} < \bar{s}) & \mathbf{1}(S_{i0} < \bar{s}) & 0 & \ldots & 0 \\ \mathbf{1}(S_{i3} < \bar{s}) & \mathbf{1}(S_{i2} < \bar{s}) & \mathbf{1}(S_{i1} < \bar{s}) & \mathbf{1}(S_{i0} < \bar{s}) & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \mathbf{1}(S_{i,T-2} < \bar{s}) & \mathbf{1}(S_{i,T-3} < \bar{s}) & \mathbf{1}(S_{i,T-4} < \bar{s}) & \mathbf{1}(S_{i,T-5} < \bar{s}) & \ldots & \mathbf{1}(S_{i,T-1-\bar{m}} < \bar{s}) \end{bmatrix}_{(T-3)*\bar{m}} . \tag{36}
$$

The term $\sum_{j=1}^{\bar{m}} Z_{iys}[*j]$ is a *vector* that is constructed by summing up all columns in Z_{iys}. In particular, for individual i, the vector will be nonzero if her stock S_{it} is smaller than $\bar{s}$ in at least one time period in the sample. Since the vector represents the extra instrument that is needed for estimating β_{S1}, the number of overidentifying restrictions remains unchanged.

Extended Mazzolari Model. The extended Mazzolari model (e.g., Column 5 in Table II) contains lagged welfare participation (i.e., $y_{i,t-1}, y_{i,t-2}, y_{i,t-3}$) as additional regressors. The model uses the same instruments as the original specification to account for the endogeneity of the stock S_{it}. It uses moment conditions in equation (17) to construct instruments for lagged welfare participation.

The model is estimated by GMM, where the full instrument matrix for individual i is

$$Z_i = \left[\begin{array}{c|c} Z_{iy} & 0 \\ \hline 0 & Z_{i\tilde{x}} \end{array}\right], \tag{37}$$

where

$$Z_{i\tilde{x}} = \begin{bmatrix} E_i \mathbf{1}\{Z_{ij0} < H_{i0}\} & E_i \mathbf{1}\{Z_{ij0} < H_{i0}\}\frac{Z_{ij0}}{H_{i0}} & E_i & \boldsymbol{X}_{i0} \\ E_i \mathbf{1}\{Z_{ij1} < H_{i1}\} & E_i \mathbf{1}\{Z_{ij1} < H_{i1}\}\frac{Z_{ij1}}{H_{i1}} & E_i & \boldsymbol{X}_{i1} \\ \vdots & \vdots & \vdots & \vdots \\ E_i \mathbf{1}\{Z_{ij,T-1} < H_{i,T-1}\} & E_i \mathbf{1}\{Z_{ij,T-1} < H_{i,T-1}\}\frac{Z_{ij,T-1}}{H_{i,T-1}} & E_i & \boldsymbol{X}_{i,T-1} \end{bmatrix}_{T\times(K_x+3)} . \tag{38}$$

Extended GM Model. The extended GM model in Table III and Table VI contains lagged welfare participation (i.e., $y_{i,t-1}, y_{i,t-2}, y_{i,t-3}$) as additional regressors. It uses moment conditions in equation (17) to construct instruments for lagged welfare participation. The model is estimated by GMM, and the full instrument matrix for individual i is

$$Z_i = \left[\begin{array}{c|c} Z_{iy} & 0 \\ \hline 0 & Z_{ix} \end{array}\right], \tag{39}$$

where Z_{iy} and Z_{ix} are defined in equations (22) and (23), respectively.

References

Anderson, T. and C. Hsiao (1982): "Formulation and Estimation of Dynamic Models using Panel Data," *Journal of Econometrics*, 18, 47–82.

Arellano, M. and S. 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.

Arellano, M. and O. Bover (1995): "Another Look at the Instrumental Variable Estimation of Error-Components Models," *Journal of Econometrics*, 68, 29–51.

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

Bitler, M. P. and H. W. Hoynes (2010): "The State of the Social Safety Net in the Post-Welfare Reform Era," *Brookings Papers on Economic Activity*, Fall 71–127.

Blank, R. M. (2001): "What Causes Public Assistance Caseloads to Grow?" *Journal of Human Resources*, 36, 85–118.

Bloom, D., J. Kemple, P. Morris, S. Scrivener, N. Verma, and R. Hendra (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.

Chan, M. K. (2013): "A Dynamic Model of Welfare Reform," *Econometrica*, 81, 941–1001.

Fang, H. and M. Keane (2004): "Assessing the Impact of Welfare Reform on Single Mothers," *Brookings Papers on Economic Activity*, 35, 1–116.

Grogger, J. (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, J. and C. Michalopoulos (2003): "Welfare Dynamics Under Time Limits," *Journal of Political Economy*, 111, 530–554.

Haider, S. and J. Klerman (2005): "Dynamic Properties of the Welfare Caseload," *Labour Economics*, 12, 629–648.

Keane, M. and K. Wolpin (2002): "Estimating Welfare Effects Consistent with Forward-Looking Behavior, Part II: Empirical Results," *Journal of Human Resources*, 37, 600–622.

——— (2010): "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*, 51, 851–892.

Klerman, J. and S. Haider (2004): "A Stock-Flow Analysis of the Welfare Caseload," *Journal of Human Resources*, 39, 865–885.

Kline, P. and M. Tartari (2014): "Bounding the Labor Supply Responses to a Randomized Welfare Experiment: A Revealed Preference Approach," Working Paper.

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

Moffitt, R. (1992): "Incentive Effects of the U.S. Welfare System: A Review," *Journal of Economic Literature*, 30, 1–61.

Ribar, D. C., M. Edelhoch, and Q. Liu (2008): "Watching the Clocks: The Role of Food Stamp Recertification and TANF Time Limits in Caseload Dynamics," *Journal of Human Resources*, 43, 208–239.

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

Ziliak, J. P. (1997): "Efficient Estimation with Panel Data when Instruments are Predetermined: An Empirical Comparison of Moment-Condition Estimators," *Journal of Business and Economic Statistics*, 15, 419–431.

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

TABLE I

SUMMARY STATISTICS OF THE BASELINE AND DISADVANTAGED SAMPLES[a]

	Baseline Sample		Disadvantaged Sample	
	Control	Treatment	Control	Treatment
Age of youngest child (years)	5.50	5.43	4.93	4.56
	(4.52)	(4.34)	(4.26)	(4.01)
Race (black=1, %)	38.87	41.70	64.18	66.03
	(48.78)	(49.34)	(47.90)	(47.39)
Age (years)	30.59	30.34	28.97	28.31
	(7.01)	(7.46)	(7.72)	(7.44)
Years of education	11.31	11.23	10.98	10.73
	(1.50)	(1.56)	(1.46)	(1.60)
Number of children	1.82	1.78	2.15	2.20
	(0.90)	(0.91)	(1.05)	(1.08)
Welfare participation rate (%)				
Full sample period	37.55	36.06	59.42	66.36
At month 5	56.54	56.49	74.07	81.13
At month 11	39.73	39.19	63.61	70.84
At month 17	31.41	27.34	54.15	61.06
At month 23	23.43	18.55	47.13	51.11
Welfare participation rate (%) by cumulative months of welfare use since random assignment:[b]				
0	18.62	19.76	36.08	45.18
1 to 3	45.67	47.86	61.58	67.25
4 to 6	40.57	42.82	61.83	65.84
7 to 9	54.00	46.12	62.21	73.63
10 to 12	57.77	53.20	72.44	75.52
13 to 15	69.79	56.20	78.13	78.86
16 to 18	67.72	69.22	81.03	81.45
19 to 21	81.55	67.22	87.86	87.81
22 to 23	95.00	89.13	95.05	92.27

a Sample means are reported. Standard deviations are given in parentheses.

b Summary statistics are constructed using observations from all months during the sample period.

TABLE II

ESTIMATION RESULTS OF THE BASELINE MODEL AND EXISTING REGRESSION APPROACHES (BASELINE SAMPLE)[a]

Variable (parameter in parentheses)	GM Model (Data from First 6 Months Only) (1)	GM Model (2)	Extended GM Model (with Stock Effects) (3)	Mazzolari Model (4)	Extended Mazzolari Model (with Lagged y)[b] (5)	Baseline GMM Model (6)
y_{t-1} (α_1)					0.681 ***	0.660 ***
					(0.016)	(0.020)
y_{t-2} (α_2)					0.118 ***	0.116 ***
					(0.013)	(0.013)
y_{t-3} (α_3)					0.008	0.006
					(0.010)	(0.010)
FTP dummy×I{A<16}×(A-16) (β_A)	0.003	0.014 ***				
	(0.003)	(0.002)				
FTP dummy×I{S<H}×(A-16) (β_A)			0.015 ***			0.003 *
			(0.002)			(0.002)
FTP dummy×I{S<H} (β_0)			-0.203 ***	-0.184 ***	-0.014	0.016
			(0.021)	(0.026)	(0.013)	(0.023)
FTP dummy×I{S<H}×(S-24) (β_S)			-0.044 ***			0.006 *
			(0.001)			(0.003)
FTP dummy×I{S<H}×(S/H) (β_1)				0.424 ***	0.018	
				(0.053)	(0.024)	
FTP dummy (β_E)	0.062 **	0.103 ***	0.041 **	0.091 ***	0.005	0.034 *
	(0.031)	(0.015)	(0.020)	(0.024)	(0.013)	(0.018)
R-squared	0.02	0.06	0.19	0.06	0.74	0.74
Sargan test of overidentifying restrictions[c]	-	-	-	-	25.37 (df=21)	23.36 (df=20)
Arellano-Bond test of serial correlation[d]	-	-	-	-	0.392	0.116

a Other regressors include age, race, years of schooling, number of children, age of the youngest child, a dummy variable that equals one if the youngest child is under age 3, the interaction between the dummy variable above with the FTP dummy, calendar year indicators, calendar month indicators, and a quartic time trend. Standard errors are given in parentheses. *, Significant at the 10 percent level; **, significant at the 5 percent level; *, significant at the 1 percent level.

b The instruments for estimating key policy parameters (β_0, β_1, and β_E) are identical to column 4. See Appendix 2 for further details.

c The Sargan statistic follows a chi-square distribution with degree of freedom (df) given in parentheses.

d The m2 statistic (Arellano and Bond (1991)) follows an asymptotic standard normal distribution.

TABLE III

PREDICTED POLICY EFFECTS FROM SELECTED MODELS, BASELINE SAMPLE[a]

	Effect of Time Limit			Effect of FTP other than Time Limit[b]
	GM	Stock	Total	
Average policy effects within sample period:				
Baseline model	-9.0	-5.6	-14.6	+10.7
GM model	-13.8	+0.0	-13.8	+10.3
Mazzolari model	-10.8	-0.7	-11.4	+9.1
Extended GM model (with lagged Y)[c]	-9.4	+0.0	-9.4	+8.1
Policy effects by month:				
Baseline model				
At month 5	-5.8	-0.1	-5.9	+7.3
At month 11	-9.9	-4.3	-14.1	+12.1
At month 23	-10.5	-12.5	-23.0	+11.3
Extended GM model (with lagged Y)[c]				
At month 5	-5.8	+0.0	-5.8	+4.7
At month 11	-10.3	+0.0	-10.3	+8.7
At month 23	-11.3	+0.0	-11.3	+10.4

a The policy effects on welfare participation are expressed in percentage points.

b The effect pertains to the treatment status coefficient (β_E) only.

c Details of model estimation are given in the Appendix.

TABLE IV

ALTERNATIVE SPECIFICATIONS OF THE BASELINE MODEL (BASELINE SAMPLE)[a]

Variable (parameter in parentheses)	Control for Work Experience[b] (1)	Piecewise Linear Stock Effect[b] (2)	Fewer Lags of Y (3)	Fewer Instruments[c] (4)	More Instruments[c] (5)
y_{t-1} (α_1)	0.638 ***	0.664 ***	0.731 ***	0.661 ***	0.672 ***
	(0.021)	(0.019)	(0.016)	(0.020)	(0.019)
y_{t-2} (α_2)	0.105 ***	0.118 ***		0.112 ***	0.120 ***
	(0.013)	(0.013)		(0.013)	(0.012)
y_{t-3} (α_3)	-0.004	0.006		0.003	0.009
	(0.010)	(0.010)		(0.010)	(0.009)
FTP dummy×I{S<H}× (A-16) (β_A)	0.003	0.003 **	0.003	0.003 *	0.003 *
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
FTP dummy×I{S<H} (β_0)	0.029	0.003	0.019	0.013	0.011
	(0.027)	(0.023)	(0.027)	(0.023)	(0.021)
FTP dummy×I{S<H}× (S-24) (β_S)	0.005 *	0.001	0.007 ***	0.005 *	0.004 *
	(0.003)	(0.004)	(0.002)	(0.003)	(0.003)
FTP dummy×I{S<H}× min{S-6,0} (β_{S1})		0.033 *			
		(0.018)			
FTP dummy (β_E)	0.022	0.026	0.041 *	0.036 **	0.031 *
	(0.024)	(0.018)	(0.022)	(0.018)	(0.017)
Work experience (γ_w)	-0.007 **				
	(0.003)				
$Work_t$ (γ_0)	-0.131 ***				
	(0.041)				
$Work_{t-1}$ (γ_1)	-0.039				
	(0.028)				
$Work_{t-2}$ (γ_2)	0.033 *				
	(0.012)				
$Work_{t-3}$ (γ_3)	-0.023 *				
	(0.013)				
R-squared	0.77	0.75	0.72	0.74	0.74
Sargan test of over-identifying restrictions[d]	29.95	23.87	80.85 ***	10.70	36.20
	(df=27)	(df=20)	(df=18)	(df=8)	(df=32)
Arellano-Bond test of serial correlation[e]	-0.51	0.02	5.98 ***	0.45	0.18

a Other regressors include age, race, years of schooling, number of children, age of the youngest child, a dummy variable that equals one if the youngest child is under age 3, the interaction between the dummy variable above with the FTP dummy, calendar year indicators, calendar month indicators, and a quartic time trend. Standard errors are given in parentheses. *, Significant at the 10 percent level; **, significant at the 5 percent level; *, significant at the 1 percent level.

b Details of model estimation are given in the Appendix.

c Columns 4 and 5 involve six and eighteen lagged dependent variables in instrument construction, respectively.

d The Sargan statistic follows a chi-square distribution with degree of freedom (df) given in parentheses.

e The m2 statistic (Arellano and Bond (1991)) follows an asymptotic standard normal distribution.

TABLE V

ESTIMATION RESULTS OF SELECTED MODELS (DISADVANTAGED SAMPLE)[a]

Variable (parameter in parentheses)	GM Model (1)	Baseline GMM Model (2)	Control for Work Experience[b] (3)
y_{t-1} (α_1)		0.583 ***	0.580 ***
		(0.023)	(0.024)
y_{t-2} (α_2)		0.165 ***	0.162 ***
		(0.015)	(0.016)
y_{t-3} (α_3)		-0.024 **	-0.019 *
		(0.012)	(0.012)
FTP dummy×I{A<15}×(A-15) (β_A)	0.012 ***		
	(0.002)		
FTP dummy×I{S<H}×(A-15) (β_A)		0.004 *	0.004 *
		(0.002)	(0.002)
FTP dummy×I{S<H} (β_0)		-0.004	-0.007
		(0.029)	(0.033)
FTP dummy×I{S<H}× (S-36) (β_S)		0.003 **	0.003 **
		(0.002)	(0.002)
FTP dummy (β_E)	0.158 ***	0.080 ***	0.088 **
	(0.016)	(0.030)	(0.036)
Work experience (γ_w)			0.002
			(0.004)
$Work_t$ (γ_0)			-0.021
			(0.039)
$Work_{t-1}$ (γ_1)			0.010
			(0.025)
$Work_{t-2}$ (γ_2)			0.007
			(0.012)
$Work_{t-3}$ (γ_3)			0.007
			(0.014)
R-squared	0.08	0.73	0.74
Sargan test of over-identifying restrictions[c]	-	25.97	31.50
	-	(df=20)	(df=27)
Arellano-Bond test of serial correlation[d]	-	-1.03	-0.506

a Other regressors include age, race, years of schooling, number of children, age of the youngest child, a dummy variable that equals one if the youngest child is under age 3, the interaction between the dummy variable above with the FTP dummy, calendar year indicators, calendar month indicators, and a quartic time trend. Standard errors are given in parentheses. *, Significant at the 10 percent level; **, significant at the 5 percent level; *, significant at the 1 percent level.

b Details of model estimation are given in the Appendix.

c The Sargan statistic follows a chi-square distribution with degree of freedom (df) given in parentheses.

d The m2 statistic (Arellano and Bond (1991)) follows an asymptotic standard normal distribution.

TABLE VI

PREDICTED POLICY EFFECTS FROM SELECTED MODELS (DISADVANTAGED SAMPLE)[a]

	Effect of Time Limit			Effect of FTP (other than Time Limit)[b]
	GM	Stock	Total	
Average policy effects within sample period:				
Baseline model	-10.9	-9.0	-19.9	+23.0
GM model	-11.1	+0.0	-11.1	+15.8
Mazzolari model	-9.7	-2.2	-11.9	+17.5
Extended GM model (with lagged Y)[c]	-9.9	+0.0	-9.9	+18.1
Policy effects by month:				
Baseline model				
At month 5	-7.9	-3.0	-10.9	+16.6
At month 11	-12.1	-8.2	-20.2	+25.6
At month 23	-12.0	-15.8	-27.8	+25.0
Extended GM model (with lagged Y)[c]				
At month 5	-6.9	+0.0	-6.9	+11.9
At month 11	-10.9	+0.0	-10.9	+19.6
At month 23	-11.0	+0.0	-11.0	+21.5

a The policy effects on welfare participation are expressed in percentage points.

b The effect pertains to the treatment status coefficient (β_E) only.

c Details of model estimation are given in the Appendix.

APPENDIX TABLE A1

DETAILED ESTIMATES OF THE BASELINE MODEL[a]

Variable (parameter in parentheses)	Baseline Sample (1)		Disadvataged Sample (2)	
y_{t-1} (α_1)	0.660	***	0.584	***
	(0.020)		(0.023)	
y_{t-2} (α_2)	0.116	***	0.165	***
	(0.013)		(0.015)	
y_{t-3} (α_3)	0.006		-0.023	**
	(0.010)		(0.012)	
FTP dummy×I{S<H}×{A-16} (β_A)	0.003	*	0.004	*
	(0.002)		(0.002)	
FTP dummy×I{S<H} (β_0)	0.016		-0.004	
	(0.023)		(0.029)	
FTP dummy×I{S<H}×(S-24) (β_S)	0.006	*	0.003	**
	(0.003)		(0.002)	
FTP dummy (β_E)	0.034	*	0.081	***
	(0.018)		(0.030)	
Age of youngest child	-0.004	***	-0.004	***
	(0.001)		(0.002)	
Race (black=1)	0.028	***	0.037	***
	(0.007)		(0.009)	
Age	0.001		0.001	*
	(0.001)		(0.001)	
Years of education	-0.005	***	-0.007	***
	(0.002)		(0.002)	
Number of children	0.000		0.015	***
	(0.003)		(0.004)	
Age of youngest child under 3 years	0.001		-0.017	
	(0.010)		(0.011)	
FTP dummy × age of youngest child under 3	0.010		0.004	
	(0.014)		(0.016)	
Time trend** 2	0.039		0.065	
	(0.050)		(0.056)	
Time trend** 3	-0.014		-0.025	
	(0.018)		(0.020)	
Time trend** 4	0.002		0.003	
	(0.002)		(0.003)	
Intercept	-0.029		-0.045	
	(0.193)		(0.215)	

a Estimates for calendar year and calendar month indicators are not reported. Standard errors are given in parentheses. *, Significant at the 10 percent level; **, significant at the 5 percent level; *, significant at the 1 percent level.