Understanding Female Welfare Participation 1991-2003

Gideon Magnus
Department of Economics
University of Chicago
gfmagnus@uchicago.edu

June 15, 2008

PRELIMINARY AND INCOMPLETE. PLEASE DO NOT CITE.

Abstract

In this paper I investigate the determinants of female welfare participation from 1991 to 2003. I estimate the parameters of a large dimensional dynamic distribution using Bayesian Model Averaging with data from from the Survey of Income and Program Participation (SIPP). The main difference between this method and methods commonly used in the welfare literature is that I do not impose a strong prior (model specification) on the distribution of the data. There were some significant changes to U.S. welfare policy during the 1990s. Most importantly, in 1996 congress passed the Personal Responsibilities and Work Responsibilities Act (PRWORA), which gave individual states much greater authority to shape policy.

My preliminary results are as follows. 1) The aggregate female welfare participation rate dropped from about 3.9% in 1991 to 1.6% in 2003. 2) The cause of this decline was the *combination* of welfare time limits and work requirements. There is no evidence that either one of these, by themselves, had any effect. 3) Single unemployed mothers were by far the most likely group to be on welfare. In addition they formed the majority of welfare recipients. 4) The main contributors to the drop were single unemployed mothers (0.42%), single unemployed women with no children (0.30%), single employed mothers (0.23%), single employed women with no children (0.17%), and married employed mothers (0.16%).

JEL codes: C11, C33, C52, I38

Keywords: Bayesian Analysis, Evaluation of welfare programs, Model Averaging

1 Introduction

In 1935 the federal government passed the Social Security Act, which started the first federal welfare program: Aid to Families with Dependent Children (AFDC). The goal was to provide financial support to single parents, mainly widowed mothers. The number of recipients was initially low, but began to increase in the 1960s, reaching an all-time high of 5.5% of the population in 1994. About 70% of recipients were children. It became increasingly accepted that the program had 'perverse' effects. That is, the program was thought to induce a significant number of women to become eligible for assistance. For example, financial aid was only given to single parents, creating an incentive to be single, as opposed to married parents. Reforms were made in 1967, 1981, and 1988, but the biggest change came in 1996 when congress passed the Personal Responsibilities and Work Responsibilities Act (PRWORA). AFDC was renamed Temporary Assistance to Needy Families (TANF). A key feature of PRWORA was that it gave individual states a much greater degree of freedom to shape policy. In general, policies have become 'tougher', making it harder to be eligible, and including more requirements to receive welfare payments. The number of recipients dropped sharply, to around 2% of the population in 2000.

My goal in this paper is to estimate and understand the factors that determine the fraction of the female population (age 15 and older) on welfare. I will focus on the years 1991 to 2003 using data from the Survey of Income and Program Participation (SIPP), a longitudinal data set. Although this topic has been extensively researched, I hope to contribute by applying a novel statistical method: I estimate the parameters a large dimensional dynamic distribution using Bayesian Model Averaging. The main advantage of my method is that it does not impose a strong prior on the distribution of the data. This is in contrast with priors (model specifications) commonly used in the literature on welfare, which in my opinion are very dogmatic.

2 Literature review

Blank (2002), Moffit (2003), and Grogger and Karoly (2005) provide overviews of the research on welfare. Fang and Keane (2004) focus on single mothers using repeated cross sections from the Current Population Survey from 1980 to 2002. They conclude that stricter work requirements were the most important factor driving caseload declines and that changes in the Earned Income Tax Credit (EITC) were the most important for increased labor supply. The latter result had been found earlier by Meyer and Rosenbaum (2001).

Welfare policy might also affect other behavioral outcomes, such as marriage and fertility. Grogger and Bronars (2001) examine the effects of welfare benefits on marriage and fertility using twin births as an instrument. They find that higher benefit levels led white single mothers to postpone marriage and black single mothers to have an additional baby more quickly. Grogger (2002) looks at the effects of time limits on participation. He expects women with younger children to be more responsive, as a woman whose youngest child is closer to adulthood adult will become ineligible for benefits in any case. A woman with young children has an incentive to refrain from going on welfare to avoid the risk of being ineligible later on. Indeed he finds that welfare participation dropped much more for mothers with younger children.

3 Data description

The Survey of Income and Program Participation is a longitudinal data set administered by the U.S. Census Bureau. Starting in 1984, a sample of households was selected and followed for a number of years. New samples ('panels') were started every year until 1993. After that, new panels were started in 1996, 2001, and 2004. Information was collected on all household members and interviews took place every four months.

I use SIPP panels 1990, 1991, 1992, 1993, 1996, and 2001. I select all women age 15 and older. I discard observations that have missing variables and/or errors. For some questions, women need to report four answers: one for each of the preceding months. I only include the answer for the fourth (most recent) month. I suspect that these answers are the most accurate as they do not ask a respondent to think back in time. The SIPP uses a rotation system for interviews, which means that the fourth interview month refers to one of four consecutive calender months, depending on which rotation group the woman belongs to.

4 Method

Any causal argument requires specification of different 'treatment' groups. In my case, different government policy bundles are the treatments. A particular policy bundle at time t is denoted $\boldsymbol{x}_t$[1]. I assume policies have different effects on different women. I denote 'outcome' variables (at time t) by $\boldsymbol{y}_t$.

[1]For notational simplicity, a variable x (vector $\boldsymbol{x}$, matrix $\boldsymbol{X}$) can either refer to a random variable, or its realization. The distinction, when important, should be clear from the context.

These are welfare participation, employment status, marital status, number of children under 18, and education. I call $\boldsymbol{z}_t = (\text{age}_t, \text{race})$ 'external' variables. I transform all (partially) continuous variables into discrete categorical variables. Variable definitions are given below. My causal statement is that, for every possible value of $(\boldsymbol{y}_{t-1}, \boldsymbol{z}_t)$:

$$\{p(\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}, \boldsymbol{x}_t, \boldsymbol{z}_t)\}_{\boldsymbol{x}_t \in \mathcal{X}}$$

is a collection of counterfactual outcome probabilities, where $\mathcal{X}$ is the collection of policies. That is, the probability of observing outcomes $\boldsymbol{y}_t$ for a woman with characteristics $(\boldsymbol{y}_{t-1}, \boldsymbol{z}_t)$, regardless of her factual treatment, denoted $\boldsymbol{x}_t^f$, would be $p(\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}, \boldsymbol{x}_t^{cf}, \boldsymbol{z}_t)$ if her treatment had counterfactually been $\boldsymbol{x}_t^{cf}$. My primary objective is therefore to estimate the following conditional distribution:

$$p(\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}, \boldsymbol{x}_t, \boldsymbol{z}_t)$$

I assume that the functional form of this distribution is the same for all t. Based on this distribution, I can evaluate the effects of different policies. That is, I can answer the question: "How would outcomes have changed for different types of women had they faced alternative policies?". From now on, I suppress notation of $\boldsymbol{z}_t$.

4.1 Aggregate outcomes

I am interested in the distribution of outcomes in the aggregate, so I will not only need to estimate the conditional distributions, but also the 'marginal' distributions of treatments and lagged characteristics at each point in time. In other words, how many women of each type faced which treatments? The relevant distributions are:

$$p(\boldsymbol{y}_{t-1}, \boldsymbol{x}_t) \quad t = 1, 2, \ldots, T$$

I will now describe how I go about doing this. I start by estimating the initial ($t = 1$) marginal distribution: $p(\boldsymbol{y}_0, \boldsymbol{x}_1)$ in two steps: first I estimate $p(\boldsymbol{y}_0, \boldsymbol{x}_0)$ and $p(\boldsymbol{x}_1 \mid \boldsymbol{y}_0, \boldsymbol{x}_0)$. Then $p(\boldsymbol{y}_0, \boldsymbol{x}_1)$ is obtained by simply summing out $\boldsymbol{x}_0$:

$$p(\boldsymbol{y}_0, \boldsymbol{x}_1) = \sum_{\boldsymbol{x}_0} p(\boldsymbol{x}_1 \mid \boldsymbol{y}_0, \boldsymbol{x}_0) \times p(\boldsymbol{y}_0, \boldsymbol{x}_0)$$

I then combine this with the initial conditional distribution to get the initial joint distribution:

$$p(\boldsymbol{y}_1, \boldsymbol{y}_0, \boldsymbol{x}_1) = p(\boldsymbol{y}_1 \mid \boldsymbol{y}_0, \boldsymbol{x}_1) \times p(\boldsymbol{y}_0, \boldsymbol{x}_1)$$

For the next time period ($t = 2$), I again want the marginal, which is given by

$$\begin{aligned} p(\boldsymbol{y}_1, \boldsymbol{x}_2) &= \sum_{\boldsymbol{y}_0, \boldsymbol{x}_1} p(\boldsymbol{y}_1, \boldsymbol{x}_2, \boldsymbol{y}_0, \boldsymbol{x}_1) \\ &= \sum_{\boldsymbol{y}_0, \boldsymbol{x}_1} p(\boldsymbol{x}_2 \mid \boldsymbol{y}_1, \boldsymbol{y}_0, \boldsymbol{x}_1) \times p(\boldsymbol{y}_1, \boldsymbol{y}_0, \boldsymbol{x}_1) \end{aligned}$$

I make the following assumption about the distribution of policies:

Assumption: $\boldsymbol{x}_t \mid \boldsymbol{y}_{t-1}, \boldsymbol{x}_{t-1} \perp\!\!\!\perp \boldsymbol{y}_{t-2}$

This means I can simplify as follows:

$$p(\boldsymbol{y}_1, \boldsymbol{x}_2) = \sum_{\boldsymbol{x}_1} p(\boldsymbol{x}_2 \mid \boldsymbol{y}_1, \boldsymbol{x}_1) \times p(\boldsymbol{y}_1, \boldsymbol{x}_1)$$

I already have $p(\boldsymbol{y}_1, \boldsymbol{x}_1)$ from the initial joint distribution. So I will only need to estimate $p(\boldsymbol{x}_2 \mid \boldsymbol{y}_1, \boldsymbol{x}_1)$. With the $t = 2$ marginal, I can get the $t = 2$ joint:

$$p(\boldsymbol{y}_2, \boldsymbol{y}_1, \boldsymbol{x}_2) = p(\boldsymbol{y}_2 \mid \boldsymbol{y}_1, \boldsymbol{x}_2) \times p(\boldsymbol{y}_1, \boldsymbol{x}_2)$$

I do the same for $t = 3, 4, \ldots, T$. In summary, I need to estimate $p(\boldsymbol{y}_0, \boldsymbol{x}_1)$ and $\{p(\boldsymbol{x}_t \mid \boldsymbol{y}_{t-1}, \boldsymbol{x}_{t-1})\}_{t=2}^{T}$. The joint distribution at each point in time is then given by

$$p(\boldsymbol{y}_1, \boldsymbol{y}_0, \boldsymbol{x}_1) = \sum_{\boldsymbol{x}_0} p(\boldsymbol{y}_1 \mid \boldsymbol{y}_0, \boldsymbol{x}_1) \times p(\boldsymbol{x}_1 \mid \boldsymbol{y}_0, \boldsymbol{x}_0) \times p(\boldsymbol{y}_0, \boldsymbol{x}_0)$$
$$p(\boldsymbol{y}_2, \boldsymbol{y}_1, \boldsymbol{x}_2) = \sum_{\boldsymbol{x}_1} p(\boldsymbol{y}_2 \mid \boldsymbol{y}_1, \boldsymbol{x}_2) \times p(\boldsymbol{x}_2 \mid \boldsymbol{y}_1, \boldsymbol{x}_1) \times p(\boldsymbol{y}_1, \boldsymbol{x}_1)$$
$$\dots$$
$$p(\boldsymbol{y}_{T-1}, \boldsymbol{y}_{T-2}, \boldsymbol{x}_{T-1}) = \sum_{\boldsymbol{x}_{T-2}} p(\boldsymbol{y}_{T-1} \mid \boldsymbol{y}_{T-2}, \boldsymbol{x}_{T-1}) \times p(\boldsymbol{x}_{T-1} \mid \boldsymbol{y}_{T-2}, \boldsymbol{x}_{T-2}) \times p(\boldsymbol{y}_{T-2}, \boldsymbol{x}_{T-2})$$
$$p(\boldsymbol{y}_T, \boldsymbol{y}_{T-1}, \boldsymbol{x}_T) = \sum_{\boldsymbol{x}_{T-1}} p(\boldsymbol{y}_T \mid \boldsymbol{y}_{T-1}, \boldsymbol{x}_T) \times p(\boldsymbol{x}_T \mid \boldsymbol{y}_{T-1}, \boldsymbol{x}_{T-1}) \times p(\boldsymbol{y}_{T-1}, \boldsymbol{x}_{T-1})$$

4.2 Counterfactual policies

I would like to assess what the distribution of outcomes would have been if the distribution of policies had been different. To that end, I need to specify alternative policy distributions: $p^*(\boldsymbol{y}_0, \boldsymbol{x}_0)$ and $p^*(\boldsymbol{x}_t \mid \boldsymbol{y}_{t-1}, \boldsymbol{x}_{t-1})$, and calculate the resulting alternative distribution of outcomes: $\{p^*(\boldsymbol{y}_t)\}_{t=1}^T$. This is given by:

$$p^*(\boldsymbol{y}_1) = \sum_{\boldsymbol{y}_0, \boldsymbol{x}_0, \boldsymbol{x}_1} p(\boldsymbol{y}_1 \mid \boldsymbol{y}_0, \boldsymbol{x}_1) \times p^*(\boldsymbol{x}_1 \mid \boldsymbol{y}_0, \boldsymbol{x}_0) \times p^*(\boldsymbol{y}_0, \boldsymbol{x}_0)$$
$$p^*(\boldsymbol{y}_2) = \sum_{\boldsymbol{y}_1, \boldsymbol{x}_1, \boldsymbol{x}_2} p(\boldsymbol{y}_2 \mid \boldsymbol{y}_1, \boldsymbol{x}_2) \times p^*(\boldsymbol{x}_2 \mid \boldsymbol{y}_1, \boldsymbol{x}_1) \times p^*(\boldsymbol{y}_1, \boldsymbol{x}_1)$$
$$\dots$$
$$p^*(\boldsymbol{y}_{T-1}) = \sum_{\boldsymbol{y}_{T-2}, \boldsymbol{x}_{T-1}, \boldsymbol{x}_{T-2}} p(\boldsymbol{y}_{T-1} \mid \boldsymbol{y}_{T-2}, \boldsymbol{x}_{T-1}) \times p^*(\boldsymbol{x}_{T-1} \mid \boldsymbol{y}_{T-2}, \boldsymbol{x}_{T-2}) \times p^*(\boldsymbol{y}_{T-2}, \boldsymbol{x}_{T-2})$$
$$p^*(\boldsymbol{y}_T) = \sum_{\boldsymbol{y}_{T-1}, \boldsymbol{x}_T, \boldsymbol{x}_{T-1}} p(\boldsymbol{y}_T \mid \boldsymbol{y}_{T-1}, \boldsymbol{x}_T) \times p^*(\boldsymbol{x}_T \mid \boldsymbol{y}_{T-1}, \boldsymbol{x}_{T-1}) \times p^*(\boldsymbol{y}_{T-1}, \boldsymbol{x}_{T-1})$$

where

$$p^*(\boldsymbol{y}_t, \boldsymbol{x}_t) = \sum_{\boldsymbol{y}_{t-1}, \boldsymbol{x}_{t-1}} p(\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}, \boldsymbol{x}_t) \times p^*(\boldsymbol{x}_t \mid \boldsymbol{y}_{t-1}, \boldsymbol{x}_{t-1}) \times p^*(\boldsymbol{y}_{t-1}, \boldsymbol{x}_{t-1})$$

4.3 Discussion

My approach is based on the idea that variables $\boldsymbol{y}_{t-1}$ (and $\boldsymbol{z}_t$) provide an adequate summary of where a woman is in her life. Each time period, a woman chooses a new $\boldsymbol{y}$. There will be many different $\boldsymbol{y}$ to choose from: $\prod_{k=1}^{5} \dim(y_k)$. The attractiveness of different choices depends on a woman's history. Some choices will probably be (close to) impossible or extremely unattractive. The benefits of different choices might be affected by policy ($\boldsymbol{x}_t$). My argument involves several assumptions that could be criticized. I now discuss some possible objections.

4.3.1 Policy endogeneity

Policy endogeneity means that government policy is affected by the distribution of outcome variables. A simple example: Suppose there are two states, one with a high share of welfare recipients and one with a low share. The state with a high share implements a tough policy, the state with a low share doesn't. We observe a higher share on welfare in the state with the tough policy than in the state with the lenient policy. One might conclude that the effect of a tough policy was *positive*. Most likely, though, the populations in the two states are not comparable. Suppose, for example, that there are two types of women: young and old. Young women are more likely to be on welfare than old. If one state has a higher fraction of young people it will have a higher fraction on welfare. If that state implements a tough policy and the other state (with old people) implements a lenient policy, the 'young' state might still have a higher fraction on welfare. One should be comparing young women with young women and old women with old. I hope to avoid this type of problem by comparing women who are arguably very similar.

4.3.2 General equilibrium effects

General equilibrium effects arise when the probability of a woman being on welfare is affected by aggregate outcomes. For example, a woman seeing other women on welfare might be more likely to go on welfare herself (see Nechyba (2001)). Another example: the decision to work depends on the wage rate, which in turn depends on the number of women in the economy with similar skills (see Heckman, Lochner, and Taber (1999)).

4.3.3 Omission of future/lagged variables

A woman might have different outcomes based on (expectations of) future policy. Or, lagged policy might have effects even after current policy is accounted for. Women might have different responses to policy based on further lags of their characteristics, for example the effect of policy on outcomes $\boldsymbol{y}_t$ depends on $\boldsymbol{y}_{t-1}$ but also on $\boldsymbol{y}_{t-2}$.

4.3.4 Policy induced migration

Suppose women decide to move to a different state in response to welfare policy. Women might choose to move to a state that has a more lenient policy in order to receive benefits that they couldn't otherwise. This would invalidate my causal argument if the women who move are not comparable to the women who do not. I would need to include into the analysis the effects of policies in states other than the state a woman lives in. Welfare policy induced migration has been analyzed by Meyer (2000), who finds that there is a small amount of welfare induced migration.

4.4 Current versus lagged conditioning variables

It is not uncommon in the welfare literature to condition on the current values of outcome variables. That is, to look at

$$p(\text{welfare}_t \mid \boldsymbol{y}_t^{\text{other}}, \boldsymbol{x}_t)$$

where $\boldsymbol{y}_t^{\text{other}}$ contains the outcome variables excluding welfare. Suppose, however, that a 'tough' policy induces women to leave welfare but also has an effect on marriage. In that case, a woman who chooses to be married under a tough policy may not be comparable to a woman who chooses to be married under a lenient policy. In other words, current welfare and marital status are jointly determined, or 'endogenous' with respect to current policy. I am not willing to make the assumption that current outcome variables are exogenous.

4.5 Variable definitions

I use the following variable definitions:

4.5.1 Outcome variables

welfare recipiency (2)

- receiving welfare payments

- not receiving welfare payments

employment status (2)

- employed
- unemployed

marital status (2)

- single
- married

number of children under 18 (2)

- 0
- ≥ 1

education level (3)

- less than high school degree
- high school degree
- some college \ college degree

4.5.2 Policy variables

time limits (2)

- facing time limits
- not facing time limits

work requirements (2)

- facing work requirements
- not facing work requirements

4.5.3 External variables

age (6)

- 15-24, 25-34, 35-44, 45-54, 55-64, ≥ 65

race (3)

- white + other
- black
- hispanic

5 Estimation

I estimate the distribution $p(\boldsymbol{y}_t, \mid \boldsymbol{y}_{t-1}, \boldsymbol{x}_t)$ by first decomposing it into conditionals. I make one more assumption:

Assumption: $y_{jt} \mid y_{lt} \perp\!\!\!\perp y_{l,t-1}$.

In other words, once the current value of a variable is conditioned on, its lagged value adds no additional information. This gives me

$$
\begin{aligned}
p(\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}, \boldsymbol{x}_t; \boldsymbol{\Theta}) =& p(y_{5t} \mid y_{1t}, y_{2t}, y_{3t}, y_{4t}, y_{5,t-1}, \boldsymbol{x}_t; \boldsymbol{\Theta}_5) \\
&\times p(y_{4t} \mid y_{1t}, y_{2t}, y_{3t}, y_{5,t-1}, y_{4,t-1}, \boldsymbol{x}_t; \boldsymbol{\Theta}_4) \\
&\times p(y_{3t} \mid y_{1t}, y_{2t}, y_{5,t-1}, y_{4,t-1}, y_{3,t-1}, \boldsymbol{x}_t; \boldsymbol{\Theta}_4) \\
&\times p(y_{2t} \mid y_{1t}, y_{5,t-1}, y_{4,t-1}, y_{3,t-1}, y_{2,t-1}, \boldsymbol{x}_t; \boldsymbol{\Theta}_2) \\
&\times p(y_{1t} \mid \boldsymbol{y}_{t-1}, \boldsymbol{x}_t; \boldsymbol{\Theta}_1).
\end{aligned}
$$

So my problem boils down to estimating the unknown (i.e. random) parameters of these 5 conditional distributions, collected in $\boldsymbol{\Theta} := (\boldsymbol{\Theta}_5, \boldsymbol{\Theta}_4, \ldots, \boldsymbol{\Theta}_1)$. Each $\boldsymbol{\Theta}_j$ is a matrix who's columns are multinomial probability vectors for all the possible conditioning sets of variable y_{jt}. I do exactly the same (i.e. decomposing into conditionals) for the policy distributions $p(\boldsymbol{x}_t \mid \boldsymbol{y}_{t-1}, \boldsymbol{x}_{t-1})$ and the initial distribution $p(\boldsymbol{y}_0, \boldsymbol{x}_0)$.

Estimating the parameters $\boldsymbol{\Theta}_j$ requires priors. The priors I choose reflect the belief that the multinomial probability vector of each conditioning set ('type') is similar to that of certain other conditioning sets. More precisely, I believe that the probability vector for a conditioning set $\boldsymbol{a} = (a_1, a_2, \ldots, a_h)$ is similar to the probability vectors for conditioning sets that are subsets of $\boldsymbol{a}$. A simplified example: suppose I look at the conditional distribution of current welfare participation (welfare$_t$) and the conditioning set is (welfare$_{t-1}$, high school degree$_{t-1}$, no kids$_{t-1}$). I believe that the probability of being on welfare for such a type is similar to the probability of being on welfare for the following types:

- welfare$_{t-1}$
- high school degree$_{t-1}$
- no kids$_{t-1}$
- welfare$_{t-1}$ & high school degree$_{t-1}$
- welfare$_{t-1}$ & no kids$_{t-1}$
- high school degree$_{t-1}$& no kids$_{t-1}$

I operationalize this prior as follows. For each matrix of parameters $\boldsymbol{\Theta}_j$, the prior is *hierarchical*

$$p(\boldsymbol{\Theta}_j) = \sum_l p(\boldsymbol{\Theta}_j \mid \mathcal{M}_{jl})\, p(\mathcal{M}_{jl}) \quad j = 1, \ldots, k$$

where I call the collection $\{\mathcal{M}_{jl}\}$ a set of *models*. Each model imposes a *partition* of the columns of $\boldsymbol{\Theta}_j$ into different groups. The columns within each group are identical. In other words, the distribution for each member of a group is the same. The members of a group share a common (sub)set of outcomes and this holds for all groups in a partition.

The models (partitions) I consider are *classification trees*. That is, starting with some initial variable, the columns are split into groups, one group for each of the categories of the initial variable. Then, for each of these groups a new variable is chosen, and a second split is performed. Then, repeating, new splits are performed on the 2nd stage variables. However, at each node, the splitting process could be terminated. So some branches of the tree could be much more developed than others. I limit my the space of models in the following way: If there are no observations at some node, I do not consider models that involve splitting that node further. I assign equal prior probability to all models that fit these criteria.

For each model, I need a prior $p(\boldsymbol{\Theta}_j \mid \mathcal{M}_{jl})$. Each group in the partition of a model has a parameter vector (the multinomial probability). For each vector, I choose a Dirichlet($\boldsymbol{\imath}$) prior. The Dirichlet prior is convenient as it leads to closed form solutions to the model probabilities. I want the model's priors to be non-informative. The mean of the Dirichlet($\boldsymbol{\alpha}$) distribution is $\boldsymbol{\alpha}/\boldsymbol{\imath}'\boldsymbol{\alpha}$. To be non-informative, the prior mean vector should have all elements equal, i.e. Dirichlet($\alpha\boldsymbol{\imath}$). Now I need to choose α, which determines the prior variance; the higher α, the lower the variance. A low α will assign more posterior probability to models with highly dispersed multinomial probabilities. A high α will assign more posterior probability to models with very similar probabilities. Setting $\alpha = 1$ provides a middle ground.

With a hierarchical prior, the posteriors become

$$p(\boldsymbol{\Theta}_j \mid \boldsymbol{N}) = \sum_l \underbrace{p(\boldsymbol{\Theta}_j \mid \boldsymbol{N}, \mathcal{M}_{jl})}_{\text{posterior for model } l} \times \underbrace{p(\mathcal{M}_{jl} \mid \boldsymbol{N})}_{\text{posterior model prob. } l} \qquad j = 1, \ldots, 5.$$

where $\boldsymbol{N}$ denotes the observed data. So the overall posterior is a weighted average of the posteriors of all models. This technique is called *Bayesian Model Averaging*[2]

There are certain cells that have zero probability. For example, it is not possible that a woman decreases in educational attainment. I set the prior for these cells to be zero for sure. What is the model probability for a certain conditional distribution j? For each column of $\boldsymbol{\Theta}_j$, indexed $h = 1, \ldots, H$, the likelihood function is (omitting j-subscripts)

$$\frac{n_h!}{n_{h1}!n_{h2}!\ldots n_{hG_h}!}\prod_{g=1}^{G_h}\theta_{hg}^{n_{hg}}$$

[2]See Hoeting, Madigan, Raftery and Volinsky (1999) for more on Bayesian Model Averaging

where G_h is the number of cells in column h with nonzero probability. $\boldsymbol{n}_h := (n_{h1}, \ldots, n_{hG_h})'$ is the number of occurrences in each cell and $n_h := \boldsymbol{n}_h'\boldsymbol{\imath}$ is the total number of occurrences in column h. The joint likelihood of all columns is the product of the likelihoods:

$$\begin{aligned}
&\prod_{h=1}^{H}\left[\frac{n_h!}{n_{h1}!n_{h2}!\ldots n_{G_h}!}\prod_{g=1}^{G_h}\theta_{hg}^{n_{hg}}\right]\\
=&\left[\prod_{h=1}^{H}\frac{n_h!}{n_{h1}!n_{h2}!\ldots n_{hG_h}!}\right]\times\left[\prod_{h=1}^{H}\prod_{g=1}^{G_h}\theta_{hg}^{n_{hg}}\right]\\
=&c\times\prod_{h=1}^{H}\prod_{g=1}^{G_h}\theta_{hg}^{n_{hg}}
\end{aligned}$$

where $c = \prod_{h=1}^{H}\frac{n_h!}{n_{h1}!n_{h2}!\ldots n_{hG_h}!}$. Suppose a model $\mathcal{M}^*$ assumes there are $H^* < H$ distinct parameter columns. For each column h I use a Dirichlet($\boldsymbol{\imath}_{G_h}$) prior. The model, by imposing equality constraints, reduces the *effective* number of parameters. The likelihood times the prior then becomes

$$c\prod_{h=1}^{H^*}\frac{1}{B(\boldsymbol{\imath}_{G_h})}\prod_{g=1}^{G_h}\theta_{hg}^{n_{hg}}$$

where $B(\boldsymbol{\alpha}) := \frac{\prod_j \Gamma(\alpha_j)}{\Gamma(\sum_j \alpha_j)}$ is the multinomial beta function. The posterior model probability, when the prior probability on each model is equal, is given by:

$$p(\mathcal{M}_j \mid \boldsymbol{N}) = \frac{p(\boldsymbol{N} \mid \mathcal{M}_j)}{\sum_l p(\boldsymbol{N} \mid \mathcal{M}_l)}$$

where $p(\boldsymbol{N} \mid \mathcal{M}_j) = \int p(\boldsymbol{N} \mid \boldsymbol{\Theta}, \mathcal{M}_j)p(\boldsymbol{\Theta} \mid \mathcal{M}_j)\,\mathrm{d}\boldsymbol{\Theta}$ is the marginal density of the data for model j, which is also the normalizing constant of the posterior (for model j). The normalizing constant is the likelihood times the prior divided by the posterior. The posterior is a product of Dirichlet distributions:

$$p(\boldsymbol{\Theta}_j \mid \{\boldsymbol{n}_h\}_{h=1}^{H^*}, \mathcal{M}^*) = \prod_{h=1}^{H^*}\frac{1}{B(\boldsymbol{n}_h + \boldsymbol{\imath}_{G_h})}\prod_{g=1}^{G_h}\theta_{hg}^{n_{hg}}$$

Using the fact that $\Gamma(x) = (x-1)!$ if x is an integer we get the following normalizing constant:

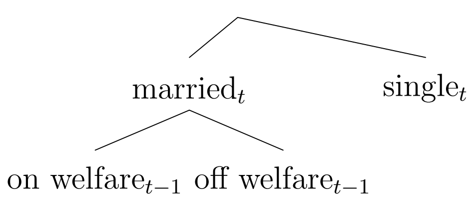
$$c\prod_{h=1}^{H^*}\frac{B(\boldsymbol{n}_h+\boldsymbol{\imath})}{B(\boldsymbol{\imath})} = c\prod_{h=1}^{H^*}\frac{\frac{\prod_{g=1}^{G_h}\Gamma(n_{hg}+1)}{\Gamma(\sum_{g=1}^{G_h}(n_{hg}+1))}}{\frac{\prod_{g=1}^{G_h}\Gamma(1)}{\Gamma(\sum_{g=1}^{G_h}1)}} = c\prod_{h=1}^{H^*}\frac{(G_h-1)!}{(n_h+G_h-1)!}\prod_{g=1}^{G_h} n_{hg}!$$

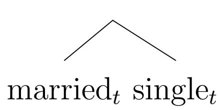
Example

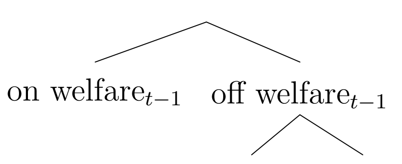
The dependent variable is current welfare participation and the conditioning variables are lagged welfare participation and marital status. The parameter matrix $\boldsymbol{\Theta}$ is given by

	1 on welfare$_{t-1}$ single$_t$	**2** off welfare$_{t-1}$ single$_t$	**3** on welfare$_{t-1}$ married$_t$	**4** off welfare$_{t-1}$ married$_t$
on welfare$_t$	θ_1	θ_2	θ_3	θ_4
off welfare$_t$	$1-\theta_1$	$1-\theta_2$	$1-\theta_3$	$1-\theta_4$

In this case, there are seven models, enumerated below. Here are three examples, represented as classification trees:







Suppose I observe the following data:

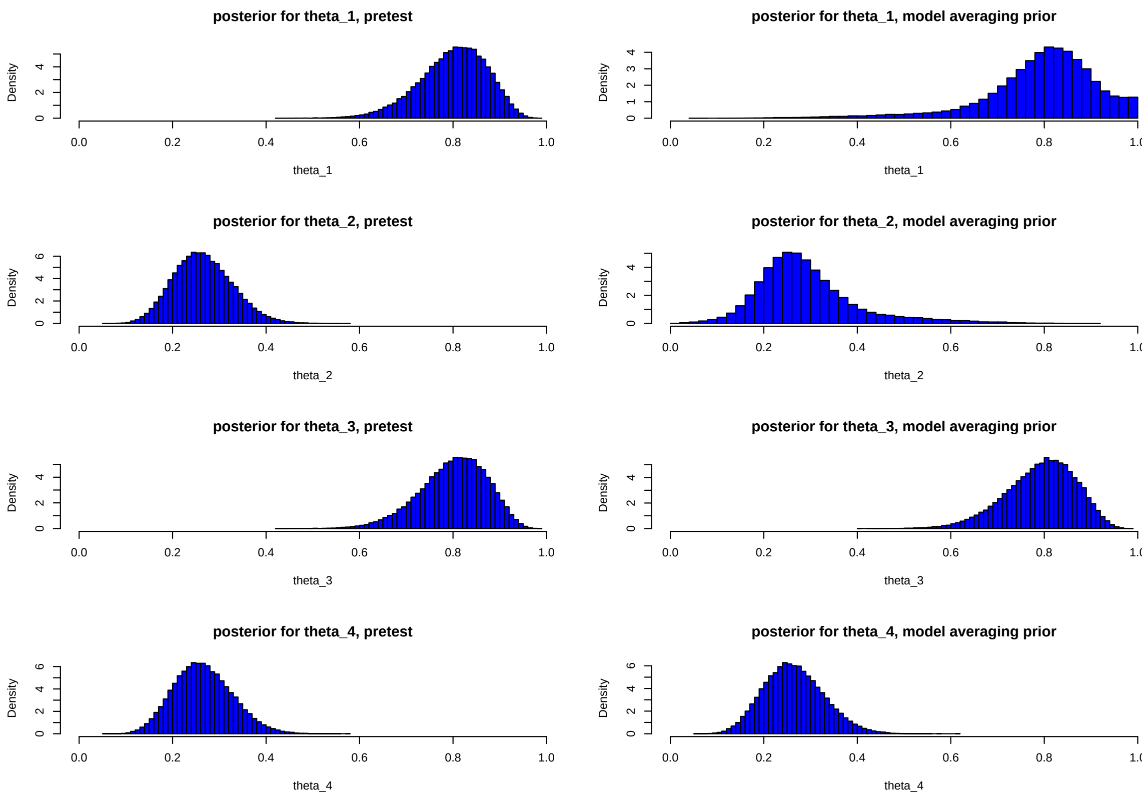
	1 on welfare$_{t-1}$ single$_t$	**2** off welfare$_{t-1}$ single$_t$	**3** on welfare$_{t-1}$ married$_t$	**4** off welfare$_{t-1}$ married$_t$
on welfare$_t$	3	2	20	10
off welfare$_t$	0	5	5	30

This gives the following posterior model probabilities:

Model 1: $\{1\};\{2\};\{3\};\{4\}$	9.98%
Model 2: $\{1,2\};\{3\};\{4\}$	2.42%
Model 3: $\{1\};\{2\};\{3,4\}$	0.0024%
Model 4: $\{1,2\};\{3,4\}$	0.00058%
Model 5: $\{1,3\};\{2\};\{4\}$	19.34%
Model 6: $\{1\};\{3\};\{2,4\}$	23.20%
Model 7: $\{1,3\};\{2,4\}$	45.03%

So model 7, that groups columns 1 and 3, as well as columns 2 and 4 together has the highest posterior probability. I could only report the posterior for this model. However, this would be inconsistent with my prior, which is that, a priori, all models are equally likely. The 'correct' posterior is different from the one that only looks at model 7. Using posteriors as priors, also known as pretesting, will generally lead to 'overconfident' predictions, as the best fitting model is used *as if* it reflects the prior that all models are equally likely. Below I show the posterior based on model 7 (left hand side) and the model averaged posterior (right hand side). In this example, we see that the posteriors based on pretesting have lower variance, especially for parameters 1 and 2.

My prior contrasts with priors commonly used in the literature on welfare. In general, papers impose what I consider a very strong prior, i.e. choice of model. The prior I use is in that sense far less dogmatic.



6 Implementation

The posterior for the parameters is a weighted average of the posteriors of all models under consideration, where the weights are given by the posterior model probabilities.

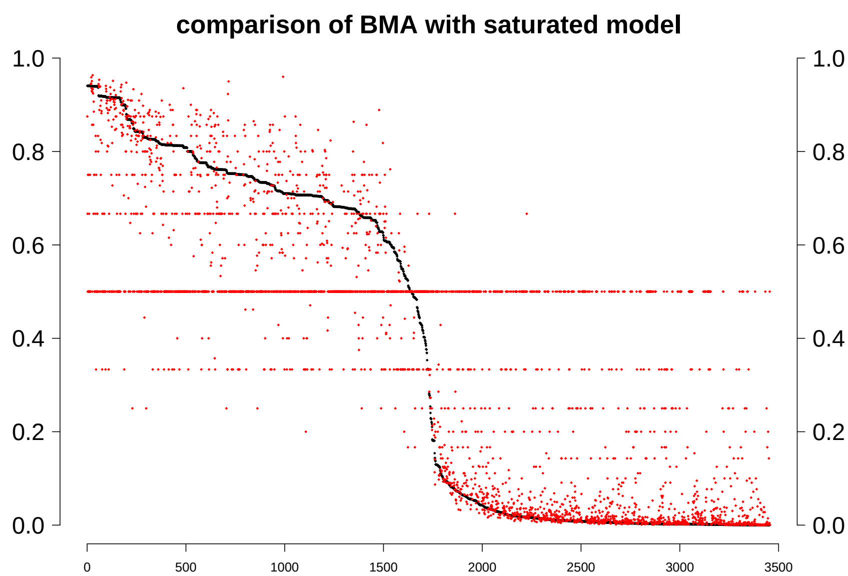
6.1 Model search algorithm

I estimate the posterior by running a model search algorithm, that seeks out high probability models while keeping a record of all models visited. I start by randomly choosing a variable and performing a split.[3] This creates a node for each of the values of the variable chosen. I calculate the log of the normalizing constant (LNC) for this model. Then for each node I draw a new variable, perform a split, and calculate the LNC for this model. If the new LNC is equal to or greater than the old LNC, the split is 'accepted' if not, it is 'rejected'. If there are no observations at a node, I don't consider any

[3] In fact, I start by picking the lagged dependent variable every time

further splits. I repeat this splitting procedure a number of times and then start again from the top. I keep track of models visited to make sure I am not double counting a model. Due to memory constraints, I exclude models that have extremely low posterior probabilities (less than $e^{-20} \approx 2 \times 10^{-9}$).

The figure below shows how the posterior obtained this way differs from the posterior of a 'saturated' model, that is, simply estimating the parameters of each conditioning set separately. Black dots are the estimated probability of being on welfare for all types of women. The types are ordered from most likely to least likely. The red dots indicate the corresponding probabilities of the saturated model. A lot of types have zero observations, which can been seen by the clustering of red dots at 0.5, which is the prior mean.[4].



6.2 Posterior generation

I simulate the full distribution over time using the estimated posterior means of $p(\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}, \boldsymbol{x}_t, \boldsymbol{z}_t)$, $p(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}, \boldsymbol{y}_{t-1}, \boldsymbol{z}_t)$, and the initial distribution $p(\boldsymbol{y}_0, \boldsymbol{x}_0)$. I set the initial period at the second third of 1991. Starting at the initial period, I generate equal sized groups of draws for women aged 15 to 80. The sampling frequency is 4 months, so for each age I have three groups (i.e. 15-1, 15-2, 15-3, 16-1, 16-2, 16-3, ... , 80-1, 80-2, 80-3). I use a group size of 500, which gives me 500*3*(80-14)=99'000 draws in total. Each period a woman shifts to the next age category and new vectors of outcomes ($\boldsymbol{y}_t$)

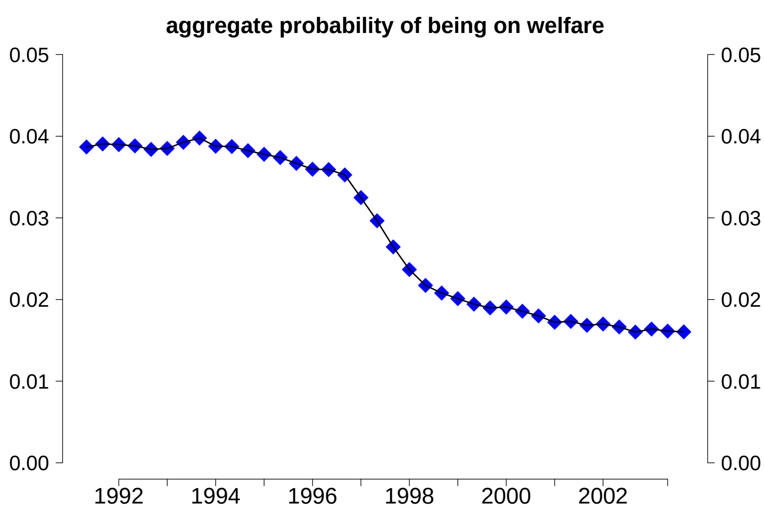
[4]A posterior of 0.5 results from observing (for a particular type) the same number of women on and off welfare, e.g. (0,0), (1,1), (2,2), etc.

and policies ($\boldsymbol{x}_t$) are drawn. Also, each period a fixed number of women die as they reach age 81. To replace these I draw new women, or 'births', i.e. 15 years olds. For this I need to estimate the distribution of 15 year olds: $p(\boldsymbol{y}_t, \boldsymbol{x}_t, \text{race} \mid \text{age}_t = 15\text{-}1)$. I do this using the exact same technique I use for the other distributions.

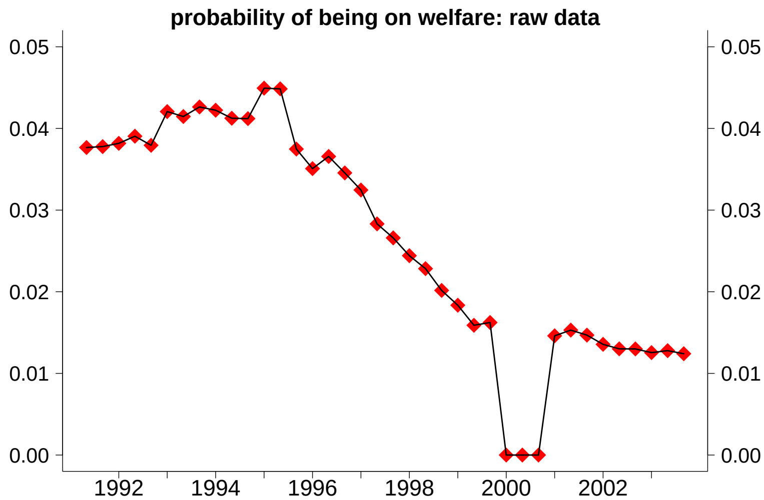
After this, I calculate posterior means at each point in time, but conditional on the 'external' variables age and race. I then multiply these conditional means by estimates of the distribution of age and race at each point in time, and sum up to get unconditional means. I use estimates of the age/race probabilities from the Statistical Abstract of the United States (SAUS). The SAUS provides yearly data, so I'm assuming here that these probabilities are fairly constant within a given year.

7 Results

I obtain the following estimate of the aggregate welfare participation rate:

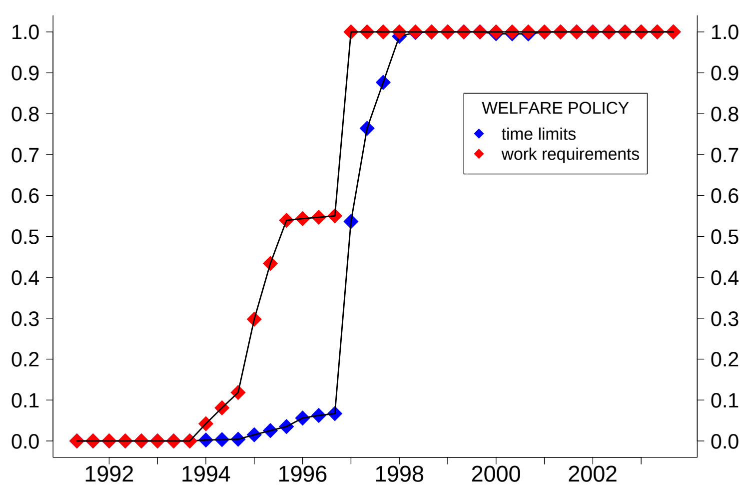


The aggregate welfare participation rate fell from 3.9% to 1.6%. As a 'reality check' I also compute, for each period, the sample average weighted by age and race:



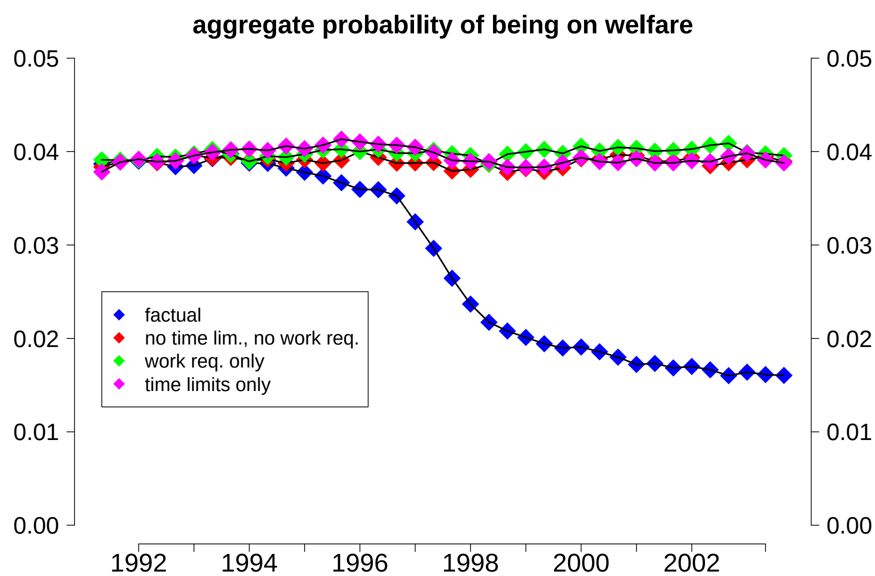
So the two are similar but by no means identical. In particular, the sample means indicate an increase and then decrease in participation between 1993 and 1996. There is no data in 2000. In the future, I will consider more complex models, i.e. models involving more policy variables and finer variables definitions. It will be interesting to see if and how my estimates change.

The next figure shows the estimated total fraction that faced time limits and work requirements over time.



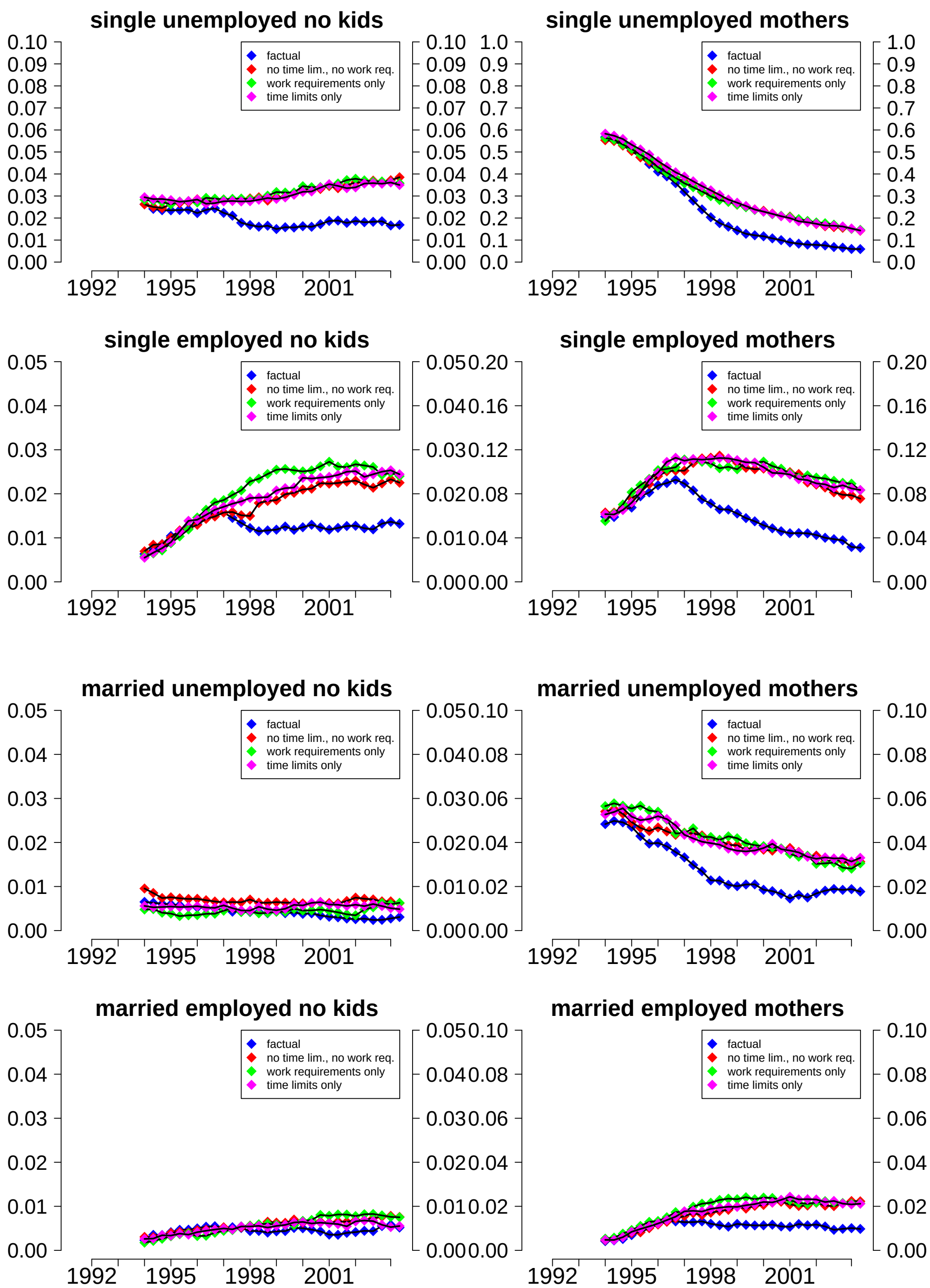
So initially there were no time limits or work requirements, but after 1994 policy starting shifting. Before 1996, states were acting under federal waivers to implement changes from the AFDC program. In 1996 PRWORA was passed and this further accelerated policy reform. By 1998 all women on welfare faced both time limits and work requirements.

I now compare the factual outcomes with three counterfactuals 1) No policy shift, that is, instead of all states eventually introducing time limits and work requirements, no states do; 2) Time limits only (no work requirements introduced); 3) Work requirements only (no time limits introduced). The next graph shows the predicted welfare participation rates under these counterfactuals.



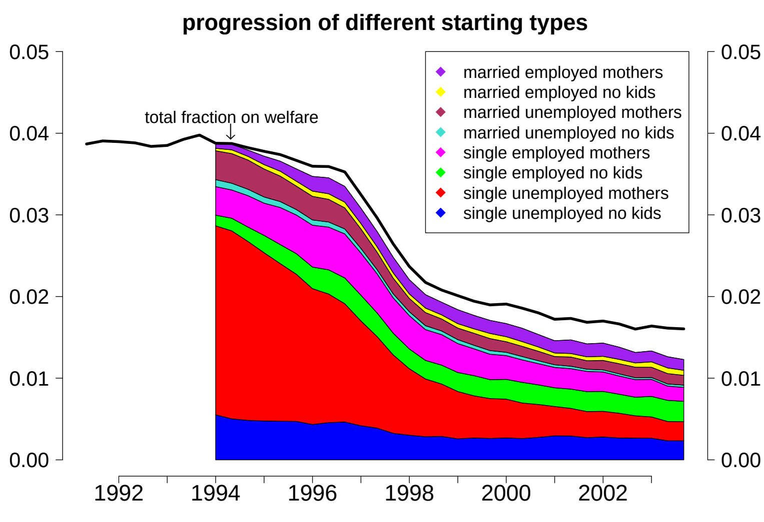
The counterfactual predictions are that the welfare participation rate remains roughly constant at 4%. There is no evidence that time limits or work requirements by themselves causes welfare participation to drop. However, the *combination* of the two did cause participation to drop. The reason I get this result is clearly the fact that the two policies were almost always implemented simultaneously.

Are there any salient characteristics of the women most affected by different policy regimes? As policy started shifting after 1994, I examine the following question: How likely were women who were of different types at the beginning of 1994 to be on welfare from 1994 to 2003 under the factual versus the counterfactual policy regimes? Note that women are characterized by their *initial* type, and may well switch types along the way. I characterize women by marital status, employment status, and number of children under 18. The following graphs show these estimates. Note the differences in scale of the vertical axes.



At the beginning of 1994, single unemployed mothers were by far the most likely group to be on welfare, with a participation rate of 60%. Single employed mothers (6%) and married unemployed mothers (4.8%) were somewhat likely to be on welfare. The least likely to be on welfare were married employed mothers (0.97%) and married employed women with no kids (0.52%). All types were more likely to be on welfare under the 'no shift in policy' counterfactual. In terms of the absolute difference in participation rate at the terminal period, the largest response was for single unemployed mothers: 8.3%, then single employed mothers at 4.5%. The least responsive by this measure were married employed women with no kids with a 0.24% difference. For all groups we see that only the combination of work requirements and time limits changed participation, with the possible exception of single employed women with no kids, who for a while appear to be *more* likely to be on welfare under a regime that involves one but no the other policy shift.

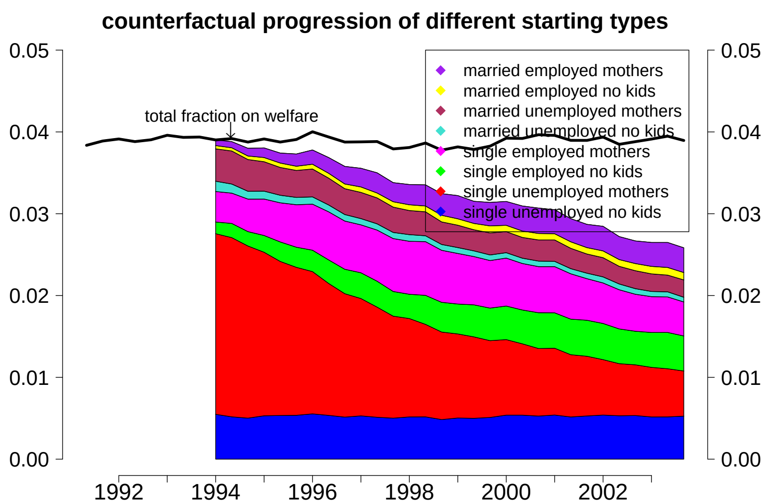
From now on, I only examine the 'no policy shift' counterfactual. I investigate how different groups contributed to the *aggregate* drop. To answer this question, I need to look at *joint* probabilities, obtained by multiplying the conditional probability of being on welfare for a particular type by the size of that type: $p(\text{welfare}_t \mid \text{type}_j) \times p(\text{type}_j)$. The factual is given in the following graph:



The sum of the probabilities does not equal the total fraction on welfare, as new women (15 year olds) enter the relevant population in each period and old women die. We see that at the beginning of 1994 a majority of welfare participants were single unemployed mothers: 60%, which is 2.3% of the total population. They are followed by single unemployed women with no kids: 14% (0.55% of the total population), then single employed mothers and married unemployed mothers, both 9.2% (0.35% of the total population).

Although single childless unemployed women were the 4th most likely group to be on welfare they did constitute, due to their size, the second largest group of welfare recipients. For each group, I compute the factual difference between the joint probability in the final period and the initial period. The group with the largest drop was single unemployed mothers: −2.1%. A distant second is single unemployed women with no kids: −0.32%.

I now look at the counterfactual:



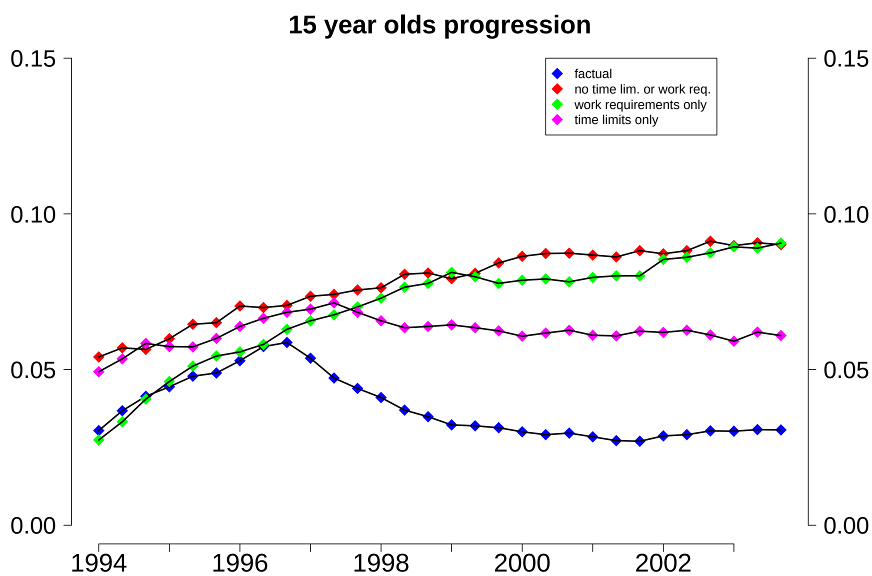
Here we see that single unemployed mothers have a smaller but still large drop: −1.65%, followed by married unemployed mothers: −0.19%. Some groups' probabilities increase: single employed women with no kids (+0.29%) and married employed mothers (+0.23%).

I now examine the difference in final period probabilities in the factual versus the counterfactual regime. The biggest differences are: single unemployed mothers (0.42%), single unemployed women with no kids (0.30%), single employed mothers (0.23%), single employed women with no kids (0.17%), and married employed mothers (0.16%).

In summary, by far the most likely group to be on welfare were single unemployed mothers, who also showed the largest response to welfare reform. They also formed the majority of welfare recipients and contributed the most to the drop in participation.

As noted before, the sum of the probabilities does not equal the total fraction on welfare. In the final time period, the total fraction on welfare is 1.6%, but I can only account for 1.2%. Under the counterfactual distribution the total fraction is 3.9% but I can only account for 2.6%. The total difference is 3.9% − 1.6% = 2.3%, of which 2.6% - 1.2% = 1.4% is accounted for. This means that 0.9% of the drop in welfare participation is due to behavioral changes by 'young' women, i.e. women who entered the population after 1994.

To get some idea of the effect of different policies on these young women, I look at the progression of women who were 15 years old in 1994 under different regimes:



From this graph we see the effects of welfare reform on this particular group. After 10 years, i.e. at age 25, the probability of being on welfare was about 9% under the 'no shift' regime and the work requirements only regime. Under the time limits only regime the probability was 6%, whereas under the factual (both policies) regime the participation rate was 3%. A more complete analysis would look at the progression of 15 year olds starting in each time period after 1994.

8 Looking ahead

I plan to examine more finely defined variables. For example, the employment variable could be split into different groups depending on how much a woman earns. The marital status variable could indicate whether a woman is separated, divorced, widowed, or never married. States imposed different types of time limits and work requirements. Also, I will include more policy variables, for example benefit levels, asset limits, and family cap (e.g. no increase in benefits after 2 children). Finally, there were other changes in policy that might affect outcomes, for example EITC and Medicaid expansions.

References

Blank, Rebecca. 2002. Evaluating Welfare Reform in the United States. *Journal of Economic Literature* 40(4):1105-66.

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

Grogger, Jeffrey. 2002. The Behavioral Effects of Welfare Time Limits. *American Economic Review* 92, no. 2: 385-89.

Grogger, Jeff, and Bronars (2001) The Effect of Welfare Payments on the Marriage and Fertility Behavior fo Unwed Mothers: Results from a Twins Experiment. *Journal of Political Economy*, vol. 109, no. 3.

Grogger, Jeffrey, and Karoly, Lynn. (2005) *Welfare Reform: effects of a decade of change*. Harvard University Press.

Heckman, J., Lochner, L., and Taber, C. "Human Capital Formation and General Equilibrium Treatment Effects: A Study of Tax and Tuition Policy," (L. Lochner and C. Taber), *Fiscal Studies*, 20(1), 25-40, (March 1999).

Hoeting, J. , Madigan, D., Raftery, A., and Volinsky, C. (1999). Bayesian Model Averaging: A Tutorial. *Statistical Science* . Vol. 14, No. 4, 382417.

Means-Tested Transfer Programs in the U.S., ed. R. Moffitt, University of Chicago Press and NBER, 2003.

Meyer, Bruce D. (2000): "Do the Poor Move to Receive Higher Welfare Benefits?" Working Paper, Northwestern University, July 1998 (revised September 2000).

Meyer, Bruce D. and Dan T. Rosenbaum. 2001. Welfare, The Earned Income Tax Credit, and the Labor Supply of Single Mothers. *Quarterly Journal of Economics*. Vol 116(3): 1063-1114. August.

Nechyba, Thomas (2001): Social Approval, Values and AFDC: A Reexamination of the Illegitimacy Debate. *Journal of Political Economy*, 109, 637-672.

Statistical Abstract of the United States, online at: http://www.census.gov/compendia/statab/