



CREAM
Centre for Research &
Analysis of Migration

Discussion Paper Series

CPD 10/18

- ▶ **The Intergenerational Transmission of Welfare Dependency**
- ▶ Monique De Haan and Ragnhild C. Schreiner

Centre for Research and Analysis of Migration
Department of Economics, University College London
Drayton House, 30 Gordon Street, London WC1H 0AX

www.cream-migration.org

The Intergenerational Transmission of Welfare Dependency*

Monique De Haan[†]

Ragnhild C. Schreiner[‡]

June 2018

Abstract

There is a strong intergenerational correlation in welfare participation, but this does not imply that parental welfare receipt induces child receipt. While there are a few quasi-experimental studies that provide estimates of the causal effect of parental welfare participation for children from marginal welfare participants, we know very little about intergenerational spillovers of welfare participation onto the children of average welfare participants. By combining rich administrative data from Norway with weak mean-monotonicity assumptions, we estimate nonparametric bounds around the average causal effect of parental welfare participation on children's welfare participation in the general population, as well as the average causal effect for children growing up in welfare-dependent families. We find that these average causal effects are considerably lower than the intergenerational correlation in welfare participation, and substantially below available local average treatment effect estimates in the literature. We further find important differences between intergenerational spillovers of disability insurance and intergenerational spillovers of financial assistance, a traditional means-tested welfare program.

Keywords: Welfare dependency, intergenerational spillovers, disability insurance, financial assistance, partial identification

JEL Classification: H55, I38, J62

*Ragnhild Schreiner acknowledges funding through a research project financed by the Norwegian Research Council (grant number 227103) and the NORFACE Welfare State Futures Programmes (GIWeS). This work is also part of Oslo Fiscal Studies at the Department of Economics, University of Oslo. Data made available by Statistics Norway have been essential for the research project.

[†]Department of Economics, University of Oslo. Also affiliated with CESifo, ESOP and Statistics Norway. Email: moniqued@econ.uio.no

[‡]Department of Economics, University College London, CReAM and the Ragnar Frisch Centre for Economic Research, Oslo. E-mail: uctpres@ucl.ac.uk.

1 Introduction

Many studies show a strong intergenerational correlation in welfare participation (see [Page \(2004\)](#) and [Black and Devereux \(2011\)](#) for overviews). It is unclear though, to what extent welfare participation of one generation induces participation by the next. Participation in an income support program is likely associated with certain (partly unobserved) characteristics that are correlated across generations, and this may create a pattern of welfare use within families. Low abilities and poor health are examples of strong predictors of welfare participation, and these characteristics are often passed on from parents to children. At the same time, there are several reasons why parental welfare participation might have a causal effect on the future welfare participation of their children. The perceived stigma associated with welfare dependence might be lower for individuals who grew up in a welfare-dependent family. Furthermore, as parents on welfare are less attached to the labor market, their ability to transfer relevant job search skills and informal job contacts to their offspring is limited, and this may create an informational disadvantage in the labor market for children from welfare-dependent families. Mirroring this, parental welfare dependence may create an informational advantage for their children when it comes to maneuvering in the social insurance system.

In order to draw lessons from intergenerational correlations in welfare participation, causation must be disentangled from the correlation that is due to shared characteristics. The literature that has attempted this is however small, and most previous studies estimate average causal effects for children of specific sub-populations of welfare participants; in most cases for the children of marginal welfare participants.¹ In case of heterogeneous effects, these local average treatment effect estimates leave us ignorant about the intergenerational spillovers of welfare programs onto the offspring of average welfare participants. If only children of marginal welfare participants are induced to become welfare participants themselves, policies that reduce the inflow into welfare programs (e.g. benefit reductions or more comprehensive screening programs) will be sufficient to curb the intergenerational spillovers. However, since these policy measures generally do not affect the average welfare participants, the existence of substantial spillovers onto the offspring of average welfare participants might require policy measures directly targeting the children, such as providing job search assistance to compensate for the lack of labor market attachment of their parents.

In this paper, we focus on two parameters: the average causal effect of parental welfare participation on children's welfare participation for the full population of children (ACE), and the average causal effect on affected children (ACAC), i.e. for children who grew up in welfare-dependent families. Identifying these parameters is not straightforward. Studies exploiting quasi-experimental variation in parental welfare participation typically identify local average treatment effects for children of particular groups of welfare participants. In order to obtain a point estimate of the average causal effect for the children of the full population of welfare participants, these studies would have to impose strong untestable assumptions – for example constant treatment effects, or assumptions regarding the comparability of marginal and average welfare participants.² To avoid having to impose these type of assumptions, we instead use

¹See for example [Antel \(1992\)](#); [Gottschalk \(1996\)](#); [Levine et al. \(1996\)](#); [Pepper \(2000\)](#); [Beaulieu et al. \(2005\)](#); [Bratberg et al. \(2014\)](#); [Dahl et al. \(2014\)](#); [Dahl and Gielen \(2018\)](#); [Hartley et al. \(2017\)](#). In Section 5 we describe the existing literature in more detail, and show how our results compare to previous findings.

²A recent paper by [Mogstad et al. \(2017\)](#) shows how to use instrumental variables to draw inferences about a wide range

a partial identification approach to estimate upper and lower bounds on the ACE and the ACAC. We exploit rich Norwegian register data, and focus on two major state-financed welfare programs; disability insurance (DI) and financial assistance (FA), that both insure against a loss or a lack of income, but that arise from different circumstances.

DI is the largest income support program in Norway (except for old age pension), and replaces approximately 66 percent of foregone earnings if a person's ability to work is significantly reduced as a result of permanent illness or injury. Almost ten percent of the population between the ages of 18 and 67 are participating in DI (NAV, 2017). The second program, FA, represents the "last layer" of income support in Norway, providing economic assistance to those without any income, or with an income that is too low to make a living. Similar to most welfare programs in Europe and the U.S., FA is strictly means tested against income and wealth. In 2017 around four percent of the population aged 18 to 67 were FA recipients (SSB, 2017). DI and FA are hence quite distinct programs, with different participants, and it can be argued that the relevance of the ACE and the ACAC differ across the two programs. We consider the average causal effect especially interesting for DI participation because of this program's broad potential coverage. Although individuals with low education and/or low income are overrepresented among the participants, individuals in all parts of the income distribution, and within all education groups may become dependent on DI benefits. As the U.S. Social Security Administration writes on its website: "Disability is a subject you may read about in the newspaper, but not think of as something that might actually happen to you. But your chances of becoming disabled are probably greater than you realize. Studies show that just over 1 in 4 of today's 20 year-olds will become disabled before reaching age 67."³ Similarly, the Norwegian Public Service Pension Fund writes on its website "Even though nobody wants to imagine being unfit for work, it is still advisable to get acquainted with the various regulations for disability pension."⁴

The average causal effect in the population is perhaps less relevant for financial assistance, as FA-dependent families are more strongly (negatively) selected, e.g. in terms of education and income. The average causal effect of parental welfare participation for individuals who grew up in a family in which at least one of the parents received welfare benefits (ACAC) is however of interest for both the FA and DI programs. As we argued above, estimates of the ACAC will, for example, inform policymakers about a potential need for compensatory policy measures targeted at the children of welfare recipients. In addition, since the participants in the DI and FA programs differ in characteristics, the comparison of the ACE's with the ACAC's will give insights into the degree of heterogeneity in the intergenerational spillovers, both between the two programs, and between children from welfare-dependent families and the general population.

of treatment parameters, including the average treatment effect (ATE), and the average treatment effect on the treated (ATT). An instrument for parental welfare participation that arguably satisfies the standard 2SLS assumptions is used by [Dahl et al. \(2014\)](#). They start out with a sample of DI applicants in Norway, whose cases are initially denied, and who are randomly assigned to appeal court judges. Next, they exploit that these judges vary in their strictness when it comes to allowing DI, and they use judge strictness as an instrument for parental DI participation. This instrument is however "undefined" for the majority of DI participants, because their DI application was immediately accepted such that they were never assigned to a (lenient or strict) judge. In practice this implies that the approach outlined in [Mogstad et al. \(2017\)](#), in combination with the judge strictness instrument, will result in wide and uninformative bounds on the ATE or ATT if one does not impose additional assumptions.

³Social Security, Benefits Planner: Disability, <https://www.ssa.gov/planners/disability/>

⁴Statens Pensjonskasse, <https://www.spk.no/en/pension/disability-pension/>

Our empirical strategy is a partial identification approach, where we impose weak mean-monotonicity assumptions that allow us to estimate informative bounds around the parameters of interest. We start by assuming that the mean potential welfare participation is non-decreasing in the actual welfare participation of the parents. In addition, we use two monotone instrumental variables (MIVs) - local labor market conditions and parental education - to tighten the bounds. When applying a traditional instrumental variables approach, the key identifying assumption is that the instrument does not have a direct effect on the potential outcomes. The identifying assumptions for *monotone* instrumental variables are less restrictive; MIV's can be weakly monotonously related to the potential outcomes. Hence, in contrast to previous studies using local labor market conditions as instrumental variables for welfare participation (see for example [Antel \(1992\)](#); [Levine et al. \(1996\)](#); [Pepper \(2000\)](#)), the possibility that local labor market conditions directly affect (or are correlated to other area characteristics that affect) the potential probability of future welfare participation of children does not violate the assumptions of our identification strategy. Similarly, by using parental education as a monotone instrument, we allow for a direct effect of parental education on children's potential welfare participation, as long as the relationship between the two variables is weakly monotonous.

Combining the above described mean-monotonicity assumptions results in informative upper bounds on the ACE and the ACAC for DI and FA. For both programs, we find that the average causal effect of parental welfare participation on children's future welfare participation is much smaller than what we would conclude on the basis of the intergenerational association. For DI we find an upper bound of 2.8 percentage points for the ACE, and 2.7 percentage points for the ACAC. For FA we find the ACE to be below 17 percentage points, and the ACAC to be below 16 percentage points, which is much larger than for DI, but still well below the intergenerational association in FA dependency. These upper bounds are not only below the estimated intergenerational associations, but also below a number of local average treatment effects reported in the previous literature. When we also add a monotone treatment response (MTR) assumption, restricting the average causal effects to be non-negative, we find informative lower bounds that indicate that parental FA participation increases children's FA participation, on average, by at least 4.2 percentage points for the population of individuals who grew up in a family dependent on FA benefits. In contrast, for DI participation the estimated lower bound, under the MTR assumption, is only slightly above zero.

Our results show that a substantial portion of the observed intergenerational correlation in welfare dependency is due to correlated characteristics. In addition, our findings indicate that there is important heterogeneity in the causal effect of parental welfare participation on children's welfare participation both between (participants in) different welfare programs, and between children of marginal welfare participants and the average member of the (treated) population.

The remainder of the paper is organized as follows. In [Section 2](#) we describe the data, the construction of the treatment and outcome variables and the two welfare programs. In [Section 3](#) we explain the partial identification approach and the identifying assumptions. The results are shown in [Section 4](#), and in [Section 5](#) we contrast our findings to the existing literature. Finally [Section 6](#) summarizes and concludes.

2 Background and Data

In this paper, we look at two welfare programs in Norway; disability insurance and financial assistance. To motivate the choice of these two programs, we begin this section by a brief description of the Norwegian welfare system. Sections 3.1 and 3.2 give more details on the two programs, and finally, Section 3.3 describes the data and the construction of the main variables.

The national welfare system in Norway covers all inhabitants who have been residents for at least twelve months. The program is state financed through payroll taxation, and includes old-age pension, unemployment benefits, sickness benefits, vocational rehabilitation programs, disability insurance and financial assistance (See Table A1 in the Appendix for an overview of the main programs in the Norwegian welfare system).⁵

DI is the largest welfare program in Norway with a participation rate of around ten percent in the population aged 18-67 (Ellingsen, 2017). High participation rates are found in many other OECD countries as well, with disability spending on average accounting for ten percent of total public spending (OECD, 2010). In addition to DI, almost all OECD countries have comprehensive minimum income programs - comparable to financial assistance - aimed at reducing poverty and providing a minimum standard of living to working-age individuals (Immerwoll, 2012). DI and FA are hence of particular interest, both because of their broad coverage within Norway, and because of their comparability with welfare programs in other countries.⁶

2.1 Disability Insurance

To qualify for DI benefits, an individual must be aged 18 to 67, have been a member of the National Insurance System (NIS) for at least three years, and have a reduced work capacity by at least 50% due to illness or injury. The typical gateway into DI is through the sick-leave program. All employees are covered, and sick-leave spells are certified by a physician. The maximum duration of sick-leave is one year. It is also possible to enter the DI program without first being on sick-leave. This is typically the case for non-employed individuals who experience a negative health-shock.

Individuals who have a reduced work capacity of at least 50% usually first receive temporary benefits for a period of 1-3 years while activation and rehabilitation programs are attempted. If rehabilitation is not successful, the individual can apply for DI benefits. After DI benefits are granted, the Social Security Administration requires no further attempts of rehabilitation, and consequently, DI benefit receipt is usually an absorbing state. Rejected DI applicants may re-apply, and five years after the initial rejection, around 40% has been granted DI benefits (Dahl et al., 2014).

During the time period of our study (1993-2014), the DI benefit payments consisted of a flat component and an earnings-dependent component. Individuals with very low, or no previous income could qualify for a minimum DI benefit.⁷ Despite the requirement of reduced health, a substantial number of

⁵In addition to these programs there are a number of benefits and subsidies available for families with (young) children. These programs are not considered to be welfare programs because they are not means-tested, nor conditional on reduced work capacity for example due to injury or illness.

⁶In addition, they are both of unlimited duration, and eligibility is not conditional on a previous labor market attachment, such as is the case with unemployment and sickness benefits.

⁷See Rege et al. (2009) for details on the benefit amounts.

DI claims follow negative employment shocks, indicating that DI participation is not only a result of poor health, but is also linked to labor market opportunities (see e.g. [Rege et al. \(2009\)](#); [Bratberg et al. \(2014\)](#) for evidence from Norway).

2.2 *Financial Assistance*

Financial assistance is a means-tested income support program that is meant as a last resort for people who otherwise have no other means of supporting themselves. In contrast to DI, FA is not an absorbing state in the sense that regular assessments of needs are carried out. In addition to eligibility, benefit amounts are determined by a case worker at the local labor and welfare administration (NAV) office. The case workers perform a discretionary assessment of needs, taking into account the savings and other wealth of the applicant (and a potential partner). The government provides guidelines for benefit amounts, based on family size, that are meant to cover expenses for basic needs such as food, clothing, communication, household appliances and hygiene, as well as expenses to cover certain leisure and social needs.⁸ The guidelines are not binding, and they do not guarantee a minimum benefit amount; this is determined by the case workers, through case-by-case assessments ([NAV, 2012](#)). Expenses related to other necessities, including housing, electricity, insurance and household items are also covered by the FA program. As these expenses can vary depending on e.g. the region of residency of the applicant, they are not included in the official guidelines, and are instead assessed by the NAV case worker.

2.3 *Data*

We use Norwegian register data covering the entire population, and with individual identifiers that allow us to link parents and children. We merge data from the Social Security Administration, containing complete records on welfare benefit receipt over the years 1993-2014, to data from Statistics Norway, containing information on individual characteristics, including age, municipality of residence, level of education and income. We create a data set consisting of children born between 1980 and 1984, such that we can observe welfare participation of their parents when the children were in their adolescence (older than 12 and younger than 18 years old; 1993 - 2001), as well as their own welfare participation between the ages 18 to 30 (1998-2014). We drop 154 individuals whose parents at birth were younger than 16, or older than 60. We also drop 17,906 children for whom we lack information on educational attainment of both parents, or the municipality of the mother when the child was aged 12.⁹ This gives us a sample of 258,452 children.

We define the two treatment variables as binary variables taking the value one if at least one of the parents participated in DI or FA respectively, during the child's adolescence. Since we are interested in welfare dependency (as opposed to receiving an income supplement during short periods of time), we define welfare participation as receiving welfare benefits for at least six months during at least one of the years when we measure participation, and the welfare benefits should be the main source of income

⁸All parents of children below the age of 18 residing in Norway receive a child allowance. This allowance is taken into account as income when the NAV case worker determines the FA amount.

⁹78% of these children have mothers born outside Norway.

during this period.¹⁰ The outcome variables are defined analogously, taking the value one if the child received DI/FA benefits for at least six months between the ages of 18 and 30.

Table 1 reports descriptive statistics for the sample of children, and shows that 2.7% of the children in the sample received DI benefits as young adults, and 6.6% received FA benefits. Around 11% of the children grew up in a family in which at least one of the parents received DI benefits, and 4% have at least one parent that received FA benefits.¹¹ Table 1 also shows that there is a positive intergenerational association in welfare participation. Children with parents on DI benefits have a 4 percentage points higher probability of being on DI benefits themselves when aged 18 to 30. For FA benefit receipt the association is even higher; children with parents that participated in FA have a 22 percentage points higher probability of receiving FA benefits when they are between 18 and 30 years old.

Table 1. Summary Statistics

<i>Child outcomes (age 18-30)</i>	<i>Mean</i>	<i>SD</i>
DI	0.027	0.162
FA	0.066	0.249
<i>Parents' characteristics</i>	<i>Mean</i>	<i>SD</i>
DI	0.106	0.308
FA	0.040	0.195
Compulsory schooling	0.387	0.487
Completed upper secondary education	0.263	0.440
Completed tertiary education	0.350	0.477
Local unemployment rate (%)	3.747	0.907
<i>Intergenerational association in welfare participation</i>	<i>OLS estimate</i>	<i>SE</i>
DI	0.040	0.001
FA	0.223	0.002
Number of children	258 452	

Note: Baseline sample of 258,452 children born between 1980 and 1984. Parental characteristics are measured when the child was aged 12. Treatment variables are measured when the child was aged 13 to 17, and outcomes when the child is aged 18 to 30.

Table 1 also shows summary statistics on the two monotone instrumental variables; parents' level of schooling, and the mean local unemployment rate. Parents' level of schooling is defined as the maximum level of schooling of the father and the mother over three levels; compulsory schooling, completed upper secondary education and completed tertiary education.¹² The local unemployment rate is defined as the

¹⁰We define welfare benefits as the main source of income if the sum of wage and business earnings does not exceed two Base Amounts. The Base Amount is a concept used within the Norwegian pension system, and was equal to 93,634 NOK (approximately 11,300 USD) in 2017. All monetary amounts in this paper are deflated to 2017 value, based on the adjustment factor used in the Norwegian pension system. The translations to US dollars are based on the average exchange rate applying in 2017, such that 1 USD=8.26 NOK.

¹¹From 1992 to 2010 individuals who had applied for DI received temporary benefits ("foreløpig uføretrygd"), while they were waiting for a decision on their DI application. The large majority of these applicants would end up on DI, and we therefore consider children of parents who received these temporary benefits as "treated". Out of the 27,393 children with parents on DI in our sample, 2.5% are defined as "treated" because their parents received these temporary benefits instead of DI benefits. Out of these, 95% have parents whose DI application was accepted.

¹²In the few cases where we lack information on one of the parents, we use the level of schooling of the other parent.

yearly average number of registered unemployed individuals in a municipality as a share of the total number of inhabitants aged 16-66 at the end of the year.¹³ To construct the MIV, we use the municipality of residence of the mother, and we take a five-year average of the municipality unemployment rate over the years 1992-1996 when the children, who are born between 1980 and 1984, were on average 12 years old.^{14,15} Section 3 will provide more information on how we use use parents' level of education and the local unemployment rate as monotone instrumental variables.

3 Empirical Strategy

In this paper we investigate whether there exist intergenerational spillovers of welfare programs. More specifically, we are interested in the causal effect of parent's welfare participation on children's welfare participation on average in the population:

$$ACE := E[W^c(1)] - E[W^c(0)] \quad (1)$$

and the average causal effect in the sub-population of children who were affected in the sense that they grew up in a welfare-dependent family:

$$ACAC := E[W^c(1) - W^c(0) | w^p = 1] \quad (2)$$

The outcome variable W^c is a binary variable taking the value one if the child participates in the welfare program when she is between 18 and 30 years old, and w^p indicates the actual welfare participation by parents. Both the ACE and the ACAC are functions of mean potential outcomes. In the next subsections we first show how to obtain bounds around these mean potential outcomes, and next we show how we can use the bounds around these mean potential outcomes to construct bounds around the ACE and ACAC.

3.1 Worst-Case Bounds on the Mean Potential Outcomes

In this section we start by showing how to obtain the so-called Worst-Case bounds (Manski, 1989) around the following two mean potential outcomes: $E[W^c(1)]$, children's mean potential welfare participation in case at least one of the parents would have participated in the welfare program, and $E[W^c(0)]$, children's mean potential welfare participation in case of none of the parents would have participated in the welfare program. Using the law of iterated expectations, we decompose these two mean potential outcomes into observed and unobserved components:

$$E[W^c(1)] = E[W^c | w^p = 1] \cdot P(w^p = 1) + E[W^c(1) | w^p = 0] \cdot P(w^p = 0) \quad (3)$$

¹³The variable measuring the local unemployment rate is part of the Local Government Dataset constructed by Fiva et al. (2017).

¹⁴About 13 percent of the fathers lived in a different municipality from the mother when the child was 12 years old. Since most children live with their mother in the case that the parents do not live together, we use the municipality of the mother to determine the local unemployment rate.

¹⁵There are many small municipalities in Norway. Variation in the unemployment rate between these small municipalities within a given year might to a large extent be due to random fluctuations and not reflect actual differences in local labor market conditions. We therefore take a five-year average of the municipality unemployment rate.

$$E[W^c(0)] = E[W^c(0)|w^p = 1] \cdot P(w^p = 1) + E[W^c|w^p = 0] \cdot P(w^p = 0) \quad (4)$$

We observe the proportion of children with and without at least one parent on welfare benefits, $P(w^p = 1)$ and $P(w^p = 0)$, as well as the shares of children receiving welfare benefits as adults within these two groups, $E[W^c|w^p = 1]$ and $E[W^c|w^p = 0]$.¹⁶ In contrast, since $W^c(1)$ is not observed for children whose parents did not receive welfare benefits, and correspondingly, $W^c(0)$ is not observed for children with at least one parent that received welfare benefits, the mean potential outcomes $E[W^c(1)|w^p = 0]$ and $E[W^c(0)|w^p = 1]$ are unobserved. Since the probability of receiving welfare benefits cannot be above zero or below one, we can replace these unobserved mean potential outcomes by zero and one to obtain the Worst-Case bounds:

$$\begin{aligned} E[W^c|w^p = 1] \cdot P(w^p = 1) &\leq E[W^c(1)] \leq E[W^c|w^p = 1] \cdot P(w^p = 1) + P(w^p = 0) \\ E[W^c|w^p = 0] \cdot P(w^p = 0) &\leq E[W^c(0)] \leq E[W^c|w^p = 0] \cdot P(w^p = 0) + P(w^p = 1) \end{aligned} \quad (5)$$

While these Worst-Case bounds are a useful starting point, they tend to be quite wide in practice. In the following subsections we therefore introduce assumptions that we will use in order to get tighter bounds.

3.2 Monotone Treatment Selection

As discussed in the Introduction, parents who receive welfare benefits are likely to be different from parents that do not receive welfare benefits. Welfare participants typically have poorer health than non-participants, and they might systematically differ from non-participants in other characteristics that are passed onto their children. Table 2 shows that there are indeed important differences in average characteristics between the parents that participate in one of the welfare programs and those that do not. Welfare participants are on average lower educated, have substantially lower earnings, are more likely to be immigrants, are less likely to be married, and are more likely to live in a high-unemployment area. Due to these differences in parental characteristics (and potentially correlated unobserved characteristics), children that grew up in a welfare-dependent family are probably more likely to become dependent on welfare benefits themselves compared to children whose parents did not participate in a welfare program, regardless of the actual welfare receipt of their parents.

Motivated by this observation, we assume that the potential probability of receiving welfare benefits is on average non-decreasing in the actual welfare participation of the parents. This monotone treatment selection (MTS) assumption is shown in Equation 6.

$$\begin{aligned} E[W^c(1)|w^p = 1] &\geq E[W^c(1)|w^p = 0] \\ E[W^c(0)|w^p = 1] &\geq E[W^c(0)|w^p = 0] \end{aligned} \quad (6)$$

In order to assess the credibility of this assumption, it is instructive to think of the case in which the MTS assumption would be violated. Consider the hypothetical situation that all parents would receive welfare benefits; the MTS assumption would be violated if children that grew up in a family without

¹⁶More precisely; the sample counterparts are observed.

welfare benefits were, on average, *strictly more likely* to participate in a welfare program as adults compared to children from welfare-dependent families. Similarly, consider the hypothetical situation that no parent would receive welfare benefits; the MTS assumption would be violated if children from welfare-dependent families were, on average, *strictly less likely* to participate in a welfare program as adults compared to children who grew up in a family without welfare benefits. These cases are, in our view, very unlikely to reflect reality, given the observed differences in characteristics between welfare participants and non-participants shown in Table 2.

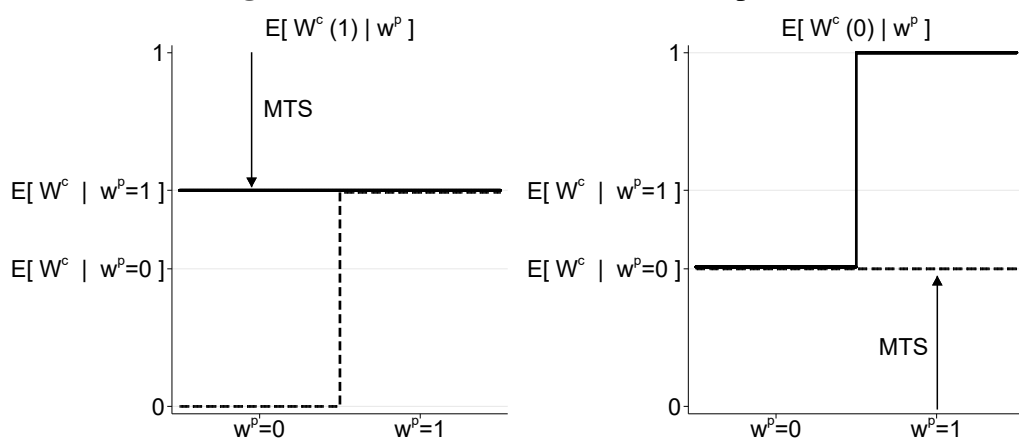
Table 2. Parental Characteristics by Welfare Dependency

	DI		FA	
	Participants ($w^p = 1$)	Non-participants ($w^p = 0$)	Participants ($w^p = 1$)	Non-participants ($w^p = 0$)
Low educated	0.53	0.30	0.71	0.30
Local unemployment rate (%)	3.87	3.74	3.95	3.74
Foreign born	0.15	0.14	0.47	0.13
Married	0.64	0.75	0.32	0.76
Earnings (\$1000)	12.47	53.45	6.12	52.71
Number of parents	30,734	515,961	14,363	532,332

Notes: The table shows descriptive statistics on the parents of the children in the baseline sample who are born between 1980-1984. Parental characteristics are measured when the child was 12 years old. Low educated is defined as having at most ten years of education. Parental earnings is the sum of wage income and business income and is converted to USD using the average exchange rate for 2017, such that 1 USD=8.26 NOK.

Figure 1 shows that under the MTS assumption, the observed mean welfare participation of children that grew up in a welfare-dependent family, $E[W^c | w^p = 1]$, can be used as an upper bound on the mean potential welfare participation of children whose parents did not participate in a welfare program in the case that their parents would have received welfare benefits, $E[W^c(1) | w^p = 0]$. Under a similar reasoning, the mean welfare participation observed among the children that grew up in a family without welfare benefits, $E[W^c | w^p = 0]$, can be used as a lower bound on $E[W^c(0) | w^p = 1]$.

Figure 1. Illustration of the MTS Assumption



Equation 7 gives the resulting MTS bounds.

$$\begin{aligned}
 E[W^c|w^p = 1] \cdot P(w^p = 1) &\leq E[W^c(1)] \leq E[W^c|w^p = 1] \\
 E[W^c|w^p = 0] &\leq E[W^c(0)] \leq E[W^c|w^p = 0] \cdot P(w^p = 0) + P(w^p = 1)
 \end{aligned}
 \tag{7}$$

3.3 Monotone Instrumental Variables

Previous papers have used local labor market conditions as instrumental variables to estimate the effect of parental participation in a welfare program on their offspring's welfare participation (see for example [Pepper 2000](#); [Levine et al. 1996](#); [Antel 1992](#)). Using local labor market conditions as an instrument implies assuming that children's participation in a welfare program is not directly affected by these labor market conditions, nor by other correlated unobserved factors. This assumption would be violated if, for example, local labor market characteristics are correlated with permanent area characteristics affecting welfare participation, or if there is serial correlation in local labor market conditions.

In the current paper, we exploit variation in local labor market conditions, however, we do not impose a mean independence assumption. Instead, we relax it to a mean monotonicity assumption, which allows for a weak monotone relation between local labor market conditions in adolescence and the mean potential welfare participation as a young adult. We use the local (municipality) mean unemployment rate, measured over the years 1992-1996, when the children were on average 12 years old, as a monotone instrumental variable (MIV). We thus assume that the potential welfare participation of individuals growing up in a high-unemployment area is not lower on average than the potential welfare participation of individuals growing up in a low-unemployment area.

Equation 8 shows the MIV assumption ([Manski and Pepper, 2000](#)), where U is the local unemployment rate.

$$E[W^c(W^p)|U = u_2] \geq E[W^c(W^p)|U = u_1] \quad \forall u_2 \geq u_1 \quad W^p \in \{0, 1\}
 \tag{8}$$

Figure 2. Illustration of the MIV Assumption

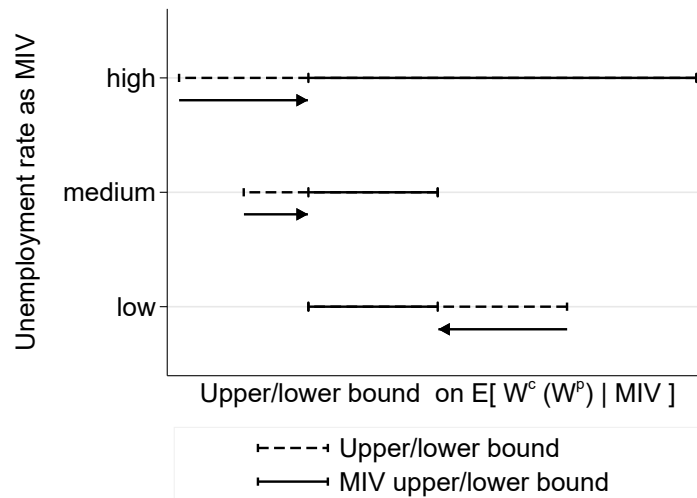


Figure 2 illustrates how the MIV assumption may give tighter bounds around the mean potential

outcomes, $E[W^c(W^P)]$. Suppose for simplicity that the local unemployment rate takes on three values: low, medium and high. Figure 2 shows illustrative lower and upper bounds within each of the three categories defined by the values of the local unemployment rate. Under the MIV assumption, the mean potential welfare participation is weakly increasing in the local unemployment rate. This implies that the mean potential welfare participation for those who grew up in a municipality with a low unemployment rate is not higher than the mean potential welfare participation for those who grew up in a municipality with a medium or high unemployment rate. The upper bound for the children growing up in a low-unemployment area can therefore be replaced by the more informative upper bound for the medium-unemployment area. Under a similar reasoning, the lower bounds for the children growing up in areas with a medium or high unemployment rate can be tightened by using the lower bound of the low-unemployment area. Equation 9 shows that aggregate MIV bounds around $E[W^c(W^P)]$ are obtained by taking the weighted average of the MIV bounds derived for each value of the local unemployment rate.

$$\begin{aligned} \sum_u P(U = u) \cdot [\max_{u_1 \leq u} LB_{E[W^c(W^P)|U=u_1]}] \\ \leq E[W^c(W^P)] \leq \\ \sum_u P(U = u) \cdot [\min_{u_2 \geq u} UB_{E[W^c(W^P)|U=u_2]}] \end{aligned} \quad W^P \in \{0, 1\} \quad (9)$$

Figure 3 shows a histogram of the mean municipality unemployment rate. The mean unemployment rate ranges from 0.9 percent to 9.4 percent, and in order to use the local unemployment rate as a monotone instrument, we need to create a categorical variable. We create quarter-percentage point categories for unemployment rates ranging from 2.25 to 4.75 percent. To make sure that we have enough observations in each category, municipalities with an unemployment rate lower than 2.25 or higher than 4.75 are combined in one top and one bottom category. This results in 12 categories, which are indicated by the vertical bars in Figure 3.¹⁷ There are on average a bit more than 25,000 observations in each category.¹⁸

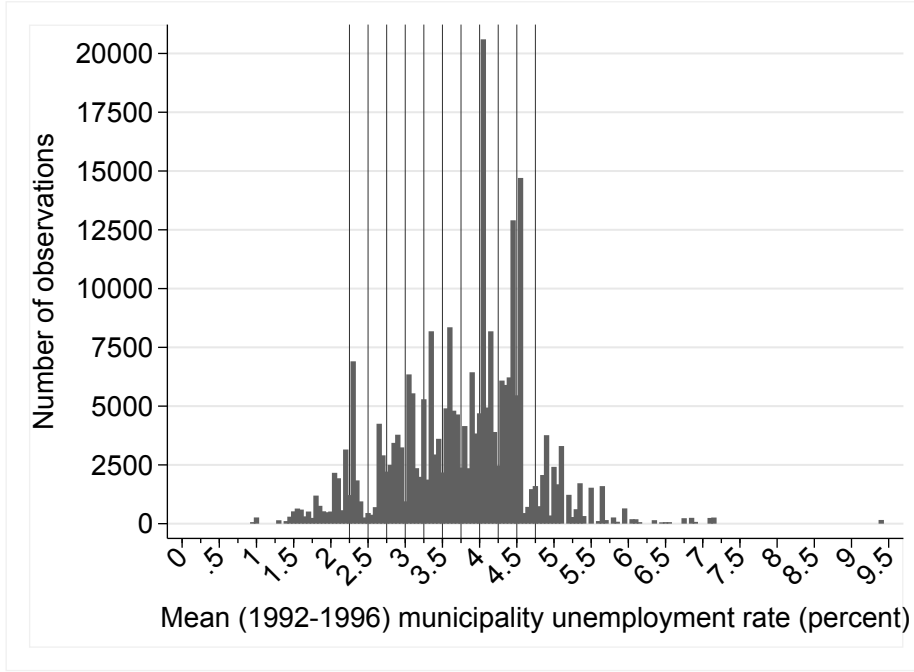
In order to assess the credibility of the MIV assumption, we use a similar reasoning as we did for the MTS assumption, and imagine the opposite case, which would imply a violation of the MIV assumption. In this case, individuals who grew up in a low-unemployment area would, on average, have a *strictly higher* potential welfare participation as adults compared to individuals that grew up in a high-unemployment area. This opposite case could arise if children growing up in high-unemployment areas invest more in their education than children from low-unemployment areas, and therefore, on average, are strictly less likely to participate in a welfare program. We find, however, that children who grew up in a high-unemployment area are not more but instead less likely to obtain a high school- or a college degree compared to children that grew up in areas with a lower unemployment rate. A regression of an indicator for obtaining a high school degree (college degree) on the mean municipality unemployment rate measured when the children were on average 12 years old gives an OLS estimate of -0.019 [s.e.=0.001] (-0.021 [s.e.=0.001]).

Another potential reason for a violation of the MIV assumption in Equation 8 is selective mobility in the sense that children from high unemployment areas systematically move to areas with better local labor market conditions. Along the same line of reasoning as before, this hypothesized scenario will

¹⁷In Section 4.3 we show that our main results are not sensitive to varying the number of categories from 8 to 14.

¹⁸The minimum number of observations in a category is 10,414 and the maximum is 40,094.

Figure 3. Histogram of Mean Municipality Unemployment, 1992-1996



only violate the MIV assumption if the selective mobility implies that children who grew up in high-unemployment areas, on average, have a strictly lower potential probability to participate in a welfare program compared to children from low-unemployment areas. Again the data indicates the contrary; children who grew up in high-unemployment areas are on average more rather than less likely to live in an area with a high unemployment rate at age 30. A regression of the unemployment rate in the municipality of residence at age 30 on the mean municipality unemployment rate, measured when the children were on average 12 years old, gives an estimate of 0.15 [s.e.=0.001]. The observed positive relation between the childhood and young adulthood unemployment rates, and the observed negative relation between the childhood unemployment rate and educational attainment make the opposite case very unlikely, and therefore strengthens the plausibility of the MIV assumption in Equation 8.

In addition to the local unemployment rate, we use parents' level of schooling as a monotone instrument. To use parents' level of schooling as an instrumental variable would imply making the assumption that the potential future welfare participation of children is mean independent of parents' level of schooling. As we argued for the local unemployment rate, mean independence is a rather strong assumption. For example, high-educated parents might invest more in the health and future labor market opportunities of their children than what low-educated parents do, and this will tend to lower the probability of their children participating in a welfare program as adults. We therefore relax this assumption, and allow for a weak monotone negative relation between the mean potential welfare participation as an adult and parents' level of schooling. Equation 10 shows the MIV assumption using parents' level of schooling (S) as a monotone instrument.

$$E[W^c(W^p)|S = s_2] \leq E[W^c(W^p)|S = s_1] \quad \forall s_2 \geq s_1, W^p \in \{0, 1\} \quad (10)$$

There exists an extensive literature documenting a strong positive relation between parents' education

and children’s educational attainment, health and labor market outcomes (see for example [Currie \(2009\)](#); [Björklund and Salvanes \(2011\)](#); [Black and Devereux \(2011\)](#) for overviews of this literature). The findings from this literature indicate that a potential relation between parents’ education and children’s welfare participation would be negative, and therefore provides strong support for the plausibility of the MIV assumption given by Equation 10.

3.4 Combining MTS and MIV Assumptions

In Section 4, we present the results imposing first only the MTS assumption, next the MTS assumption combined with one of the MIV’s, and finally the MTS assumption combined with both monotone instruments. Using both monotone instrumental variables implies imposing the following double-MIV assumption:

$$E [W^c(W^P)|U = u_1, S = s_2] \leq E [W^c(W^P)|U = u_2, S = s_1] \quad (11)$$

$$\forall u_2 \geq u_1 \text{ and } s_2 \geq s_1, W^P \in \{0, 1\}$$

This assumption states that the potential probability that a child receives welfare benefits as an adult is, on average, not higher for children with a high schooled parent ($S = s_2$) that grew up in a municipality with a low unemployment rate ($U = u_1$), compared to children who either have a lower schooled parent ($S = s_1$) and/or grew up in a municipality with a higher unemployment rate ($U = u_2$). The double-MIV assumption states nothing about the relative magnitudes of the mean potential probabilities of welfare participation when we compare children with a high schooled parent ($S = s_2$) that grew up in a municipality with a high unemployment rate ($U = u_2$) to children with a lower schooled parent ($S = s_1$) that grew up in a municipality with a low unemployment rate ($U = u_1$). The computation of the double-MIV bounds is similar to that of the single MIV-bounds, except that the maxima and minima are taken over ordered sub-samples based on paired values of the local unemployment rate and parental schooling.

To combine the MTS and the two MIV assumptions, we first compute MTS bounds within sub-samples, defined by the values of the monotone instruments, and next impose the (double) MIV assumption to tighten the bounds. This implies that the MTS assumption should hold conditional on parental education and the local unemployment rate. In Table A2 in the Appendix we show parental characteristics by welfare dependency separately for each combination of the local unemployment MIV and the parental education MIV. Within each of these sub-samples welfare participants are less likely to be married, more likely to be foreign born and have lower earnings than non-participants, which is consistent with the MTS assumption holding also conditional on the MIV’s.

3.5 Monotone Treatment Response

The potential mechanisms described in the Introduction all point to a positive causal effect of parental welfare participation on children’s welfare participation as adults. In addition, previous studies on the intergenerational transmission of welfare dependency all report positive estimates, or estimates that are not significantly different from zero.¹⁹ Hence, the controversy in the literature mostly concerns the size

¹⁹See Section 5 for more details on the estimates from the previous literature.

of the causal relation, in particular whether it is substantially greater than zero. Based on this prior of a non-negative causal effect, we therefore also show results imposing the monotone treatment response (MTR) assumption given by Equation 12.

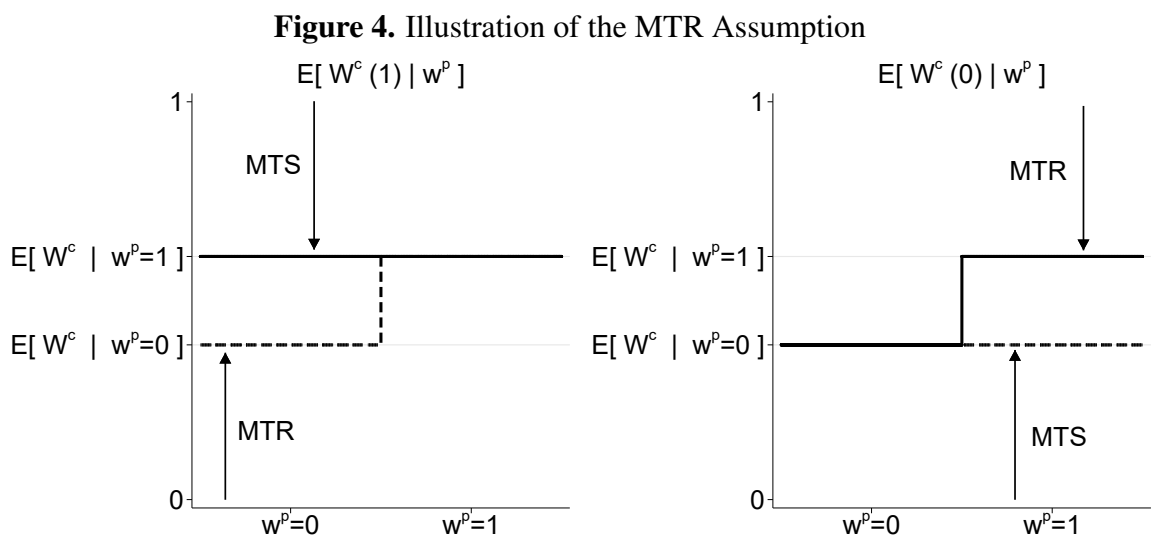
$$\begin{aligned}
 E[W^c(1)|w^p = 1] &\geq E[W^c(0)|w^p = 1] \\
 E[W^c(1)|w^p = 0] &\geq E[W^c(0)|w^p = 0]
 \end{aligned}
 \tag{12}$$

It is important to note that the inequality signs in Equation 12 are not strict, which implies that the MTR assumption does not impose a positive causal effect; rather it assumes that the causal effect is not negative.

Figure 4 illustrates how the MTR assumption can be used to tighten the bounds around the two mean potential outcomes. Under the MTR assumption, the mean potential welfare participation of children with no parents participating in a welfare program, in the case that one of their parents would have participated in a welfare program, is weakly higher than the observed mean welfare participation among these children. This implies that $E[W^c|w^p = 0]$ can be used as a lower bound on $E[W^c(1)|w^p = 0]$. The second panel of Figure 4 shows that under a similar reasoning, we can use $E[W^c|w^p = 1]$ as an upper bound on $E[W^c(0)|w^p = 1]$. Equation 13 gives the bounds on the mean potential outcomes, when imposing the combined MTR-MTS assumption.

$$\begin{aligned}
 E[W^c] &\leq E[W^c(1)] \leq E[W^c|w^p = 1] \\
 E[W^c|w^p = 0] &\leq E[W^c(0)] \leq E[W^c]
 \end{aligned}
 \tag{13}$$

To combine the MTR and MTS assumptions with the monotone instrumental variable assumptions, we first obtain MTR-MTS bounds within each sub-sample defined by the MIV's, and next impose the (double) MIV assumption to tighten the bounds.



While the MTS and MIV assumptions place restrictions on how mean potential outcomes can

vary between groups of individuals that differ in terms of characteristics, the MTR assumption places restrictions on the sign of the difference between the two mean potential outcomes for a given group of individuals. This makes the MTR assumption inherently different from the MTS and MIV assumptions. Our main analysis therefore consists of obtaining nonparametric bounds around the ACE and ACAC under different combinations of the MTS and MIV assumptions. We will thereafter continue by investigating whether we can learn more about the magnitude of the ACE and ACAC given that these average causal effects are non-negative, which implies imposing the MTR assumption.

3.6 Bounds on the ACE

In the previous subsections we showed how to use different sets of assumptions to obtain bounds around the two mean potential outcomes $E[W^c(1)]$ and $E[W^c(0)]$. In Section 4, we report bounds around the ACE, which is the difference between the two mean potential outcomes. In order to obtain an upper bound on the ACE we subtract the lower bound on $E[W^c(0)]$ from the upper bound on $E[W^c(1)]$. Similarly, subtracting the upper bound on $E[W^c(0)]$ from the lower bound on $E[W^c(1)]$ gives a lower bound on the ACE:

$$LB_{E[W^c(1)]} - UB_{E[W^c(0)]} \leq ACE \leq UB_{E[W^c(1)]} - LB_{E[W^c(0)]} , \quad (14)$$

where the exact definitions of $LB_{E[W^c(0)]}$, $UB_{E[W^c(0)]}$, $LB_{E[W^c(1)]}$ and $UB_{E[W^c(1)]}$ depend on the set of assumptions imposed.

3.7 Bounds on the ACAC

To bound the ACAC, note that this average causal effect for the group of children that grew up in a welfare-dependent family, which is defined in Equation 2, can also be written as follows:

$$ACAC = E[W^c | w^p = 1] - E[W^c(0) | w^p = 1] \quad (15)$$

While $E[W^c | w^p = 1]$, the mean welfare participation for the treated children, is observed²⁰, we do not observe $E[W^c(0) | w^p = 1]$, the mean potential welfare participation for the treated children in case their parents would not have received welfare benefits.

Equation 4 shows that by using the law of iterated expectation we can write $E[W^c(0)]$ as a weighted average of the observed mean welfare participation of the non-treated children, $E[W^c | w^p = 0]$, and the unobserved mean potential welfare participation for the treated children in case their parents would not have received welfare benefits, $E[W^c(0) | w^p = 1]$. The derived bounds on $E[W^c(0)]$, under the different combinations of the MTS, MIV and MTR assumptions, can therefore also be used as bounds on this

²⁰More precisely, the sample counterpart $E[\widehat{W^c} | w^p = 1]$ is observed.

weighted average:

$$\begin{aligned}
& LB_{E[W^c(0)]} \\
& \leq [E[W^c(0) | w^p = 1] \cdot Pr(w^p = 1) + E[W^c | w^p = 0] \cdot Pr(w^p = 0)] \leq
\end{aligned} \tag{16}$$

$$UB_{E[W^c(0)]}$$

Rewriting Equation 16 gives the following bounds around $E[W^c(0) | w^p = 1]$:

$$\begin{aligned}
& \frac{LB_{E[W^c(0)]} - E[W^c | w^p = 0] \cdot Pr(w^p = 0)}{Pr(w^p = 1)} \\
& \leq E[W^c(0) | w^p = 1] \leq
\end{aligned} \tag{17}$$

$$\frac{UB_{E[W^c(0)]} - E[W^c | w^p = 0] \cdot Pr(w^p = 0)}{Pr(w^p = 1)}$$

The ACAC can therefore be bounded as follows:

$$\begin{aligned}
& E[W^c | w^p = 1] - \frac{UB_{E[W^c(0)]} - E[W^c | w^p = 0] \cdot Pr(w^p = 0)}{Pr(w^p = 1)} \\
& \leq ACAC \leq
\end{aligned} \tag{18}$$

$$E[W^c | w^p = 1] - \frac{LB_{E[W^c(0)]} - E[W^c | w^p = 0] \cdot Pr(w^p = 0)}{Pr(w^p = 1)}$$

3.8 Estimation and Inference

Bounds estimated under the MIV assumptions might suffer from finite sample bias (Manski and Pepper (2000, 2009)). This concern is mitigated since the analysis in this paper is based on a large register data set. Nonetheless, all estimated bounds reported in Section 4 are bias-corrected using a correction method proposed by Kreider and Pepper (2007).²¹ Furthermore, to take into account sampling variability, we apply the methods from Imbens and Manski (2004) to estimate 95% confidence intervals, based on 999 bootstrap replications, around the bounds. Since there might be multiple children from one family in our data set, the sample drawn during each replication is a bootstrap sample of clusters, with clusters defined by the identifier number of the mother.²²

²¹Kreider and Pepper (2007) suggest to estimate the finite sample bias as $b\hat{ias} = (\frac{1}{K} \sum_{k=1}^K \theta_k) - \hat{\theta}$, where $\hat{\theta}$ is the initial estimate of the upper or lower bound, and θ_k is the estimate of the k^{th} bootstrap replication. The bias-corrected MIV-bounds are subsequently obtained by subtracting the estimated biases from the estimated upper and lower bounds.

²²Equation 19 gives the formula for a 95-percent confidence interval:

$$CI_{0.95} = (\hat{l}b - c_{IM} \cdot \hat{\sigma}_{lb} \quad , \quad \hat{u}b + c_{IM} \cdot \hat{\sigma}_{ub}) \tag{19}$$

where $\hat{l}b$ and $\hat{u}b$ are the estimated lower and upper bounds, and $\hat{\sigma}_{lb}$ and $\hat{\sigma}_{ub}$ are the standard errors of the estimated lower and upper bounds, obtained by 999 bootstrap replications. The parameter c_{IM} depends on the width of the bounds, and is obtained

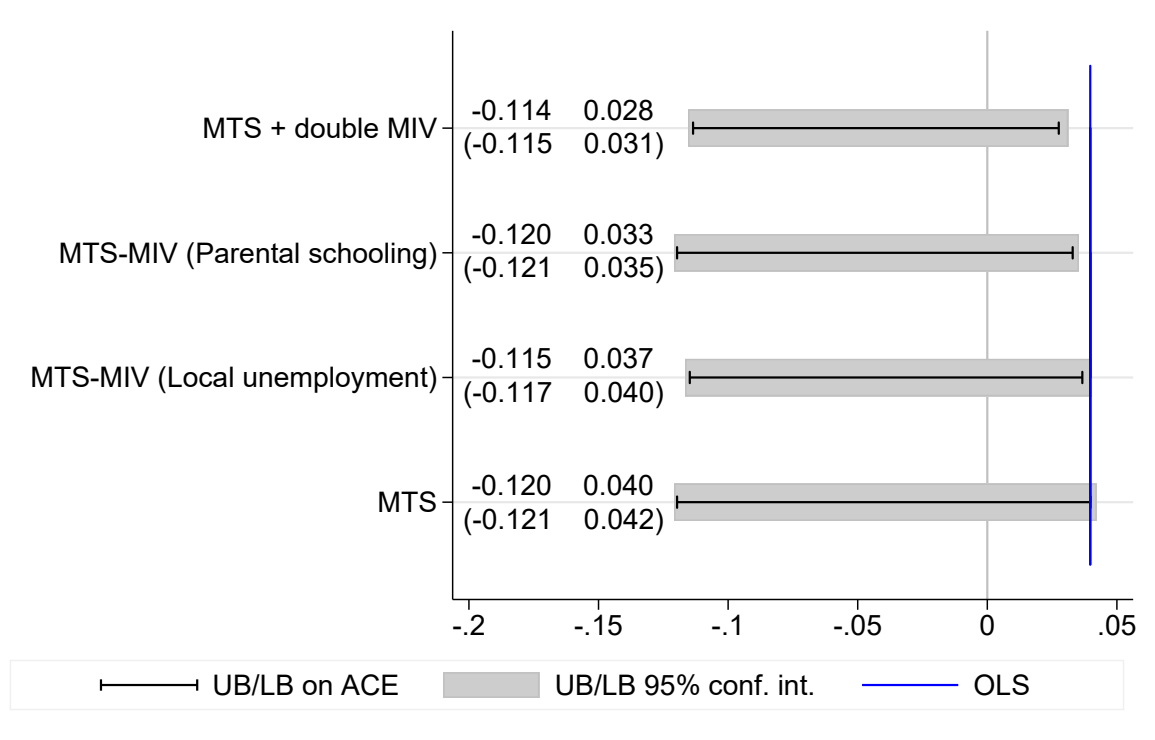
4 Results

We now turn to the empirical results, and a discussion of what can be learned about the average causal effect of parental welfare participation on children’s participation, for the general population ($E[W^c(1)] - E[W^c(0)]$), and for children who grew up with at least one parent on welfare ($E[W^c(1) - W^c(0)|w^p = 1]$), under various mean-monotonicity assumptions.

4.1 Nonparametric Bounds Around the ACE

Figures 5 and 6 show the estimated upper and lower bounds on the ACE obtained under the different sets of assumptions. We start by discussing the results for disability insurance, after which we turn to financial assistance.

Figure 5. The ACE of Parental DI participation on Children’s DI Participation



Note: Baseline sample of 258,452 children born between 1980 and 1984. The estimated bounds are corrected for finite-sample bias using the method developed by Kreider and Pepper (2007). Confidence intervals are based on the method described in Imbens and Manski (2004) and are obtained from 999 bootstrap replications. To account for correlations in welfare participation between siblings, each replication draws a cluster of siblings. There are 214,401 clusters in the sample.

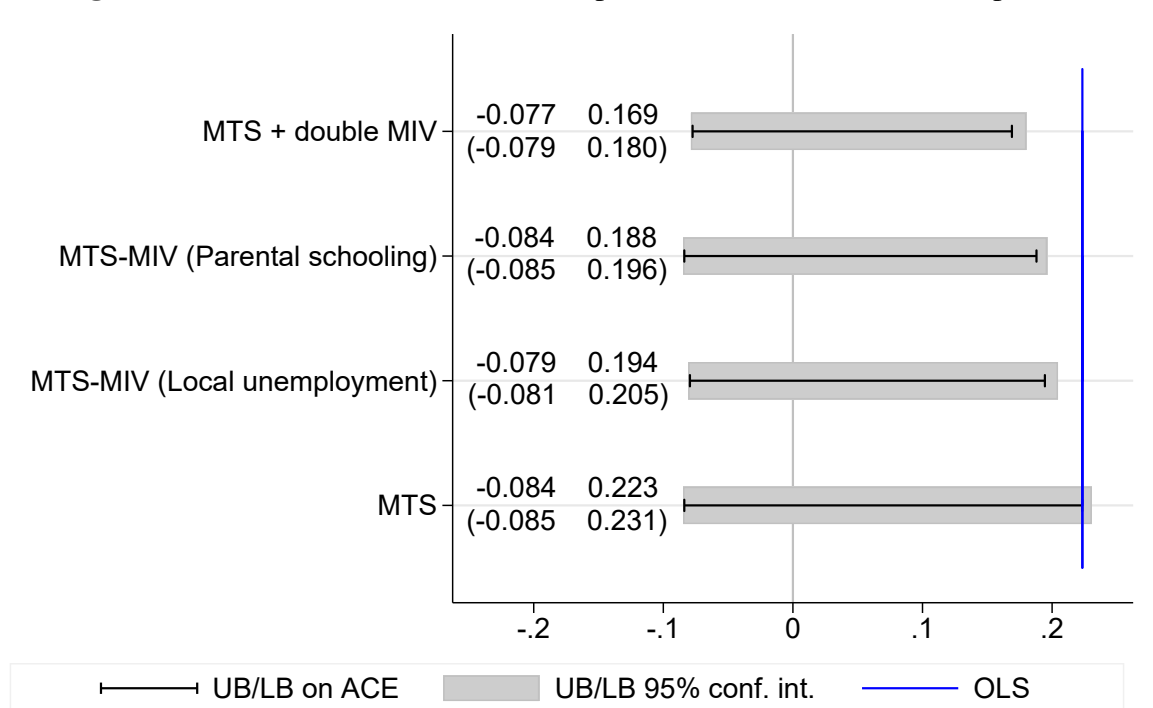
As described in Section 3, we obtain Worst-Case bounds by simply noting that the unobserved potential outcomes must lay in the interval $[0,1]$. These Worst-Case bounds for the ACE of parental DI participation on children’s DI participation equal $[-0.120 ; 0.880]$. Next, we assume that the potential DI participation is on average non-decreasing in the actual DI participation of the parents. Imposing the MTS assumption gives a more informative upper bound; parental DI participation increases, on average, by solving Equation 20.

$$\Phi \left(c_{IM} + \frac{(\hat{u}b - \hat{l}b)}{\max\{\hat{\sigma}_{lb}, \hat{\sigma}_{ub}\}} \right) - \Phi(-c_{IM}) = 0.95 \quad (20)$$

the probability that the child participates in DI by at most four percentage points. Combining the MTS assumption with either the local unemployment rate measured when the children were on average 12 years old, or the maximum level of parent’s schooling as monotone instruments tightens the upper bounds even further to respectively 3.7 and 3.3 percentage points. Finally, the tightest bounds are obtained by combining the MTS assumption with both MIV’s simultaneously, and show that the average causal effect of parental DI participation on children’s DI participation is at most 2.8 percentage points.

The OLS estimate of the relation between parents’ and children’s DI participation of 0.040 (reported in Table 1) is marked with a blue vertical line in Figure 5. A comparison of the MTS-double MIV upper bound with the OLS estimate reveals that a substantial part, at least 30 percent, of the intergenerational association in DI participation is due to correlated characteristics of parents and their children.

Figure 6. The ACE of Parental FA Participation on Children’s FA Participation



Note: Baseline sample of 258,452 children born between 1980 and 1984. The estimated bounds are corrected for finite-sample bias using the method developed by Kreider and Pepper (2007). Confidence intervals are based on the method described in Imbens and Manski (2004) and are obtained from 999 bootstrap replications. To account for correlations in welfare participation between siblings, each replication draws a cluster of siblings. There are 214,401 clusters in the sample.

For financial assistance, the Worst-Case bounds around the average causal effect equal [-0.084 ; 0.916]. As shown in Figure 6, imposing the MTS assumption reduces the upper bound on the average causal effect of parental FA participation on children’s FA participation to 22.3 percentage points. Combining the MTS assumption with the local unemployment rate as MIV reduces the upper bound further to 19.4 percentage points, and exploiting variation in bounds by parents’ level of schooling gives an upper bound of 18.8 percentage points. Finally, when combining the MTS assumption with the two monotone instruments simultaneously, we find that parental FA participation increases children’s FA participation by at most 16.9 percentage points.

The estimated upper bound on the intergenerational spillovers of FA is substantially higher than the

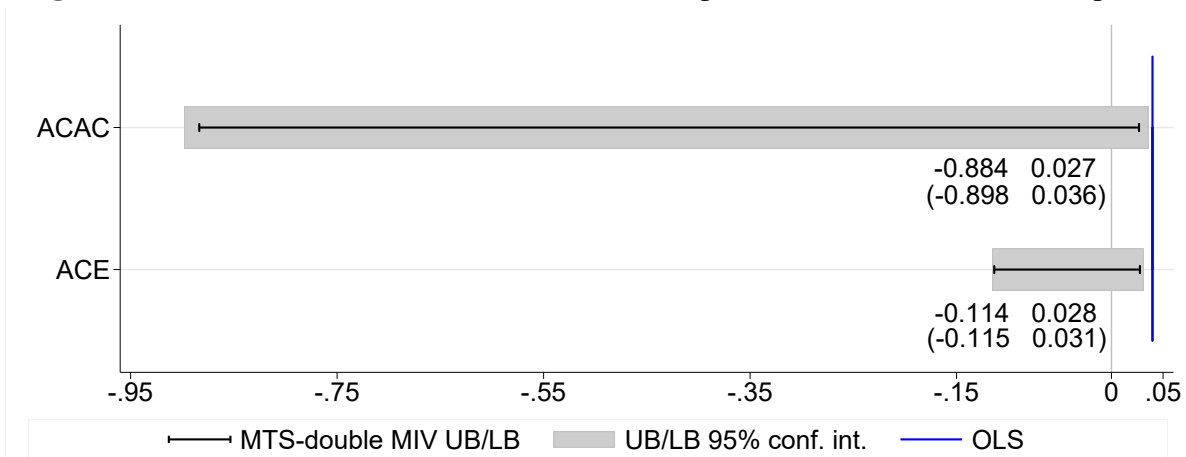
upper bound on the intergenerational spillovers of DI. However, a comparison of the OLS estimate of the relation between parents' and children's FA participation of 0.223 (reported in Table 1) with the estimated upper bound on the ACE for FA reveals that the results are qualitatively very similar to the results for DI. Also for FA a substantial part of the intergenerational association, at least 24 percent, is due to shared characteristics between parents and their children.

4.2 Nonparametric Bounds Around the ACAC

Next we turn to the average causal effect for children who grow up in a family in which at least one of the parents are dependent on welfare benefits. Figure 7 shows the estimated MTS-double MIV bounds around the ACAC as well as the MTS-double MIV bounds around the ACE for disability insurance, and Figure 8 shows the results for financial assistance.

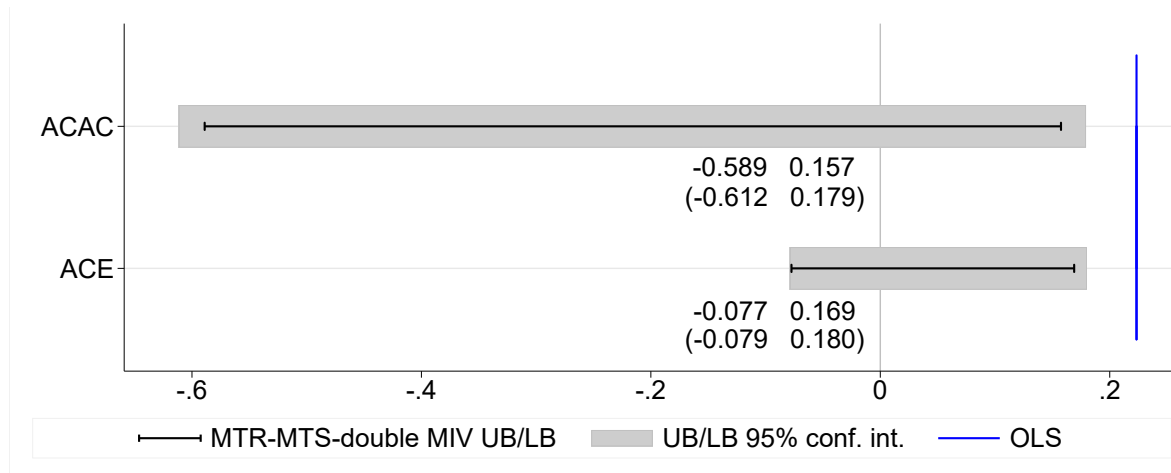
For both welfare programs, the lower bounds on the ACAC are not very informative. This can (partly) be explained by the fact that the shares of affected children in the population are rather low; 11% for DI and 4% for FA. The upper bounds are however informative, and even slightly lower than the upper bounds on the ACE. The upper bound on the ACAC for DI shows that parental DI participation increases the probability that the child participates in DI, on average, by at most 2.7 percentage points for the population of children that grew up in a DI-dependent family. For FA we find that, among the children that grew up in a family that depended on FA benefits, parental FA participation increases children's participation by at most 15.7 percentage points. For both welfare programs the 95 percent confidence intervals around the bounds on the ACAC exclude the OLS estimates (indicated by the blue vertical lines in Figures 7 and 8). Hence, not only does the intergenerational association in welfare dependency overestimate the intergenerational spillovers in the general population, it also over-estimates the intergenerational spillovers in the sub-population that grew up in welfare-dependent families.

Figure 7. The ACAC and ACE of Parental DI Participation on Children's DI Participation



Note: Baseline sample of 258,452 children born between 1980 and 1984. The estimated bounds are corrected for finite-sample bias using the method developed by Kreider and Pepper (2007). Confidence intervals are based on the method described in Imbens and Manski (2004) and are obtained from 999 bootstrap replications. To account for correlations in welfare participation between siblings, each replication draws a cluster of siblings. There are 214,401 clusters in the sample.

Figure 8. The ACAC of Parental FA Participation on Children’s FA Participation



Note: Baseline sample of 258,452 children born between 1980 and 1984. The estimated bounds are corrected for finite-sample bias using the method developed by [Kreider and Pepper \(2007\)](#). Confidence intervals are based on the method described in [Imbens and Manski \(2004\)](#) and are obtained from 999 bootstrap replications. To account for correlations in welfare participation between siblings, each replication draws a cluster of siblings. There are 214,401 clusters in the sample.

4.3 Robustness Checks

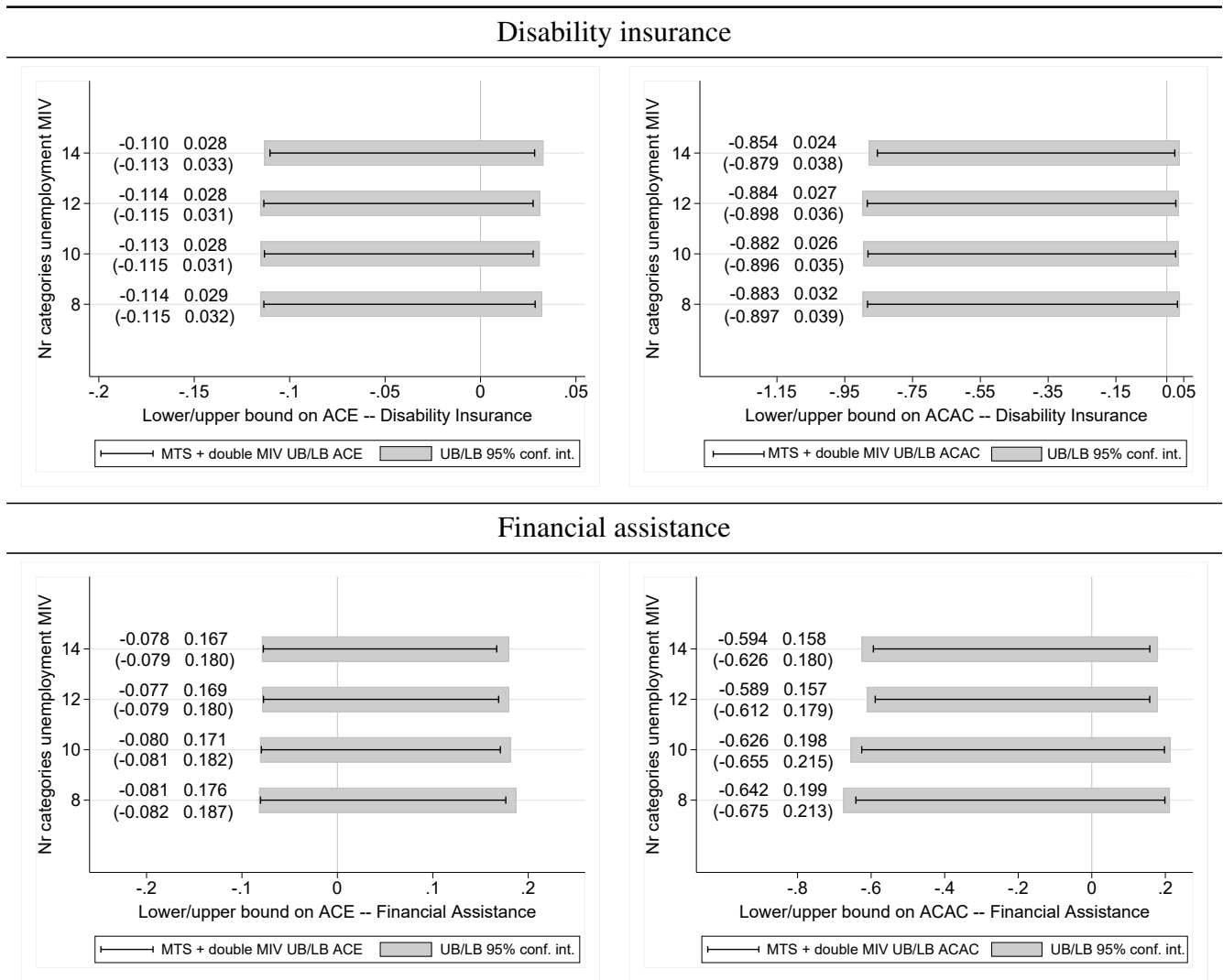
In this subsection we investigate the robustness of the main findings to varying the number of MIV categories, and to excluding the five largest municipalities. In addition we investigate whether the main findings change when we account for two programs that affected the inflow into DI and FA.

Categories of the Local Unemployment Rate

The local unemployment rate is a continuous variable, and in order to use it as a monotone instrumental variable in a nonparametric bounds analysis we have to create categories. More categories obviously gives more potential variation in upper and lower bounds between categories that can be exploited to get tighter overall bounds around the ACE and ACAC. As indicated by [Manski and Pepper \(2009\)](#) it is however important to have enough observations within each category to prevent substantial finite sample bias. Apart from this requirement, the existing literature does not provide any guidelines on how to determine categories when constructing an MIV.

As described in Section 3, the main analysis is based on a monotone instrument with 12 categories. In order to see how sensitive the results are to the choice of the number of categories, we estimate the MTS-double MIV bounds with 8, 10 and 14 categories. Figure A1 in the Appendix shows how the different numbers of categories are constructed. Figure 9 shows estimated MTS-double MIV bounds around the ACE and the ACAC when varying the number of categories from 8 to 14. Reassuringly, the results do not vary much with the number of categories.

Figure 9. Varying the Number of Categories of the Unemployment MIV,



Note: The figure shows MTS-double MIV bounds around the ACE and ACAC. Baseline sample of 258,452 children born between 1980 and 1984. The estimated bounds are corrected for finite-sample bias using the method developed by [Kreider and Pepper \(2007\)](#). Confidence intervals are based on the method described in [Imbens and Manski \(2004\)](#) and are obtained from 999 bootstrap replications. To account for correlations in welfare participation between siblings, each replication draws a cluster of siblings. There are 214,401 clusters in the sample.

Excluding the Five Largest Municipalities

The population of Norway is rather unevenly distributed over municipalities, with almost 20 percent of the children in our sample growing up in the five largest municipalities: Oslo, Bergen, Trondheim, Stavanger and Bærum.²³ Since one of the monotone instruments, the mean local unemployment rate, varies at the municipality level, our results might to a large extent be driven by these five largest municipalities. Even if this is the case, it does not necessarily imply a violation of the MIV assumption (given by Equation 8), but it would nonetheless be reassuring if the results do not change substantially when excluding these five municipalities. Row I.2 of Table 3 shows the estimated MTS-double MIV bounds on the ACE and ACAC of parental welfare participation on children’s participation when excluding the five largest municipalities. The results are very similar to the baseline results (reported in row I.1), indicating that our findings are

²³The children in our sample grew up in 428 different municipalities. Bærum is connected to Oslo, but is officially a separate municipality.

not solely driven by variation (in the unemployment rates) between the five largest municipalities.

Table 3. Robustness Checks: MTS-double MIV Bounds Around the ACE and ACAC

	OLS	ACE		ACAC		N
		LB	UB	LB	UB	
DI						
I.1: Baseline	0.040	-0.114	0.028	-0.884	0.027	258 452
		(-0.115	0.031)	(-0.898	0.036)	
I.2: Excluding five largest municipalities	0.042	-0.115	0.027	-0.898	0.030	207 497
		(-0.117	0.031)	(-0.913	0.041)	
I.3: Including Tidsbegrenset Uførestønad	0.043	-0.114	0.030	-0.879	0.027	258 452
		(-0.116	0.034)	(-0.893	0.036)	
FA						
II.1: Baseline	0.223	-0.077	0.170	-0.589	0.157	258 452
		(-0.079	0.181)	(-0.612	0.179)	
II.2: Excluding five largest municipalities	0.235	-0.077	0.191	-0.637	0.136	207 497
		(-0.077	0.207)	(-0.665	0.170)	
II.3: Including Kvalifiseringsprogrammet	0.226	-0.079	0.170	-0.582	0.158	258 452
		(-0.080	0.181)	(-0.604	0.181)	

Note: Results for a number of robustness specifications. Table report MTS-double MIV bounds. The estimated bounds are corrected for finite-sample bias using the method developed by [Kreider and Pepper \(2007\)](#). Confidence intervals are based on the method described in [Imbens and Manski \(2004\)](#) and are obtained from 999 bootstrap replications. To account for correlations in welfare participation between siblings, each replication draws a cluster of siblings.

Programs that Affected the Inflow to DI or FA

As described in Section 2, individuals who exhausted their sick-leave benefits, and who have a reduced work capacity of at least 50%, usually go onto receiving temporary benefits for a period of 1-3 years while activation and rehabilitation programs are attempted. From 1992 to 2010 these vocational and rehabilitation benefits were called *Attføring og Rehabiliteringspenger*. In 2004 another program called *Tidsbegrenset Uførestønad* was introduced alongside *Attføring og Rehabiliteringspenger*, and in 2010 the two programs were merged into one program called *Arbeidsavklaringspenger*.

Between 2004 and 2010, some DI applicants were granted a new benefit called “*Tidsbegrenset Uførestønad*”, while waiting for a decision on their DI application. Some of the participants in this program would previously have been granted benefits under the program *Attføring og Rehabiliteringspenger*. At the same time, the social security administration has estimated that around 40% of the entrants into this new program would have received disability insurance benefits in the absence of the new program ([Bragstad \(2009\)](#)). We do not consider participants in *Tidsbegrenset Uførestønad* to be DI participants in the baseline specification, but as a robustness check we treat participants in this program as regular DI participants. Since the program was implemented after the period in which we measure parental DI participation, this only affects the outcome variable. Row I.3 in Table 3 shows that the estimated MTS-double MIV bounds on the ACE and the ACAC hardly change when including participants in *Tidsbegrenset Uførestønad* in the group of DI participants.

Another benefit program affected the inflow into financial assistance. In 2007 a new comprehensive

benefit program called Kvalifiseringsprogrammet was introduced with the aim to assist hard-to-employ FA recipients in returning to the labor market. The program coupled comprehensive activation requirements with a non-means-tested benefit.²⁴ Compared to FA, participation rates in this new program are low, with around 5,700 participants in 2014 (NAV (2014)), but many of the participants were recruited from the FA program. Of the children in the main sample, 1,865 participated in Kvalifiseringsprogrammet between the ages of 18 and 30. Out of these, 73% are also classified as FA participants according to our definition. We run a robustness check where we include the remaining 27% as FA participants, and as expected, this hardly changes the results (see row II.3 of Table 3).

4.4 Monotone Treatment Response

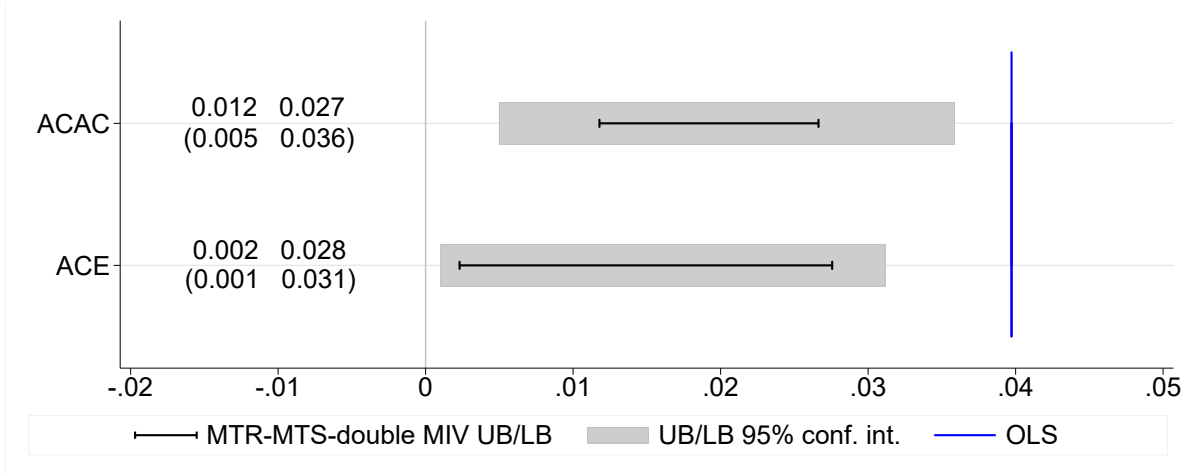
Figures 10 and 11 show estimated bounds around the ACE and ACAC when we combine the MTS and MIV assumptions with the MTR assumption. As the MTR assumption does not affect the upper bounds on the ACE and ACAC, these are identical to the upper bounds shown in Figures 7 and 8.²⁵ The lower bounds are, however, potentially affected by the MTR assumption, and it turns out that they are especially informative for the ACAC's. When combining the MTS-double MIV assumptions with the MTR assumption, we find that parental DI participation increases children's DI participation on average by at least 1.2 percentage points, and at most 2.7 percentage points among those that grew up in a DI-dependent family. For FA the average causal effect of parental FA participation on children's FA participation is at least 4.2, and at most 15.7 percentage points among those that grew up in a FA-dependent family.

The estimated lower bound on the ACAC for FA is substantially higher than the estimated upper bound on the ACAC for DI, which indicates that there are important differences between the intergenerational spillovers of disability insurance and the intergenerational spillovers of financial assistance. These differences could be due to differences in program characteristics, but an alternative explanation is that the participants differ between the two welfare programs. Table 2 shows that participants in FA indeed differ in terms of characteristics from DI participants; they are lower educated, more likely to be foreign born, less likely to be married, more likely to live in a high-unemployment area, and they have substantially lower earnings. Differences in characteristics across participants in different welfare programs, in combination with heterogeneity in the causal effect of parents' welfare participation on children's welfare participation, could therefore be an explanation for the differences in intergenerational spillovers between the two welfare programs.

²⁴See Markussen and Røed (2016) for more details on the program.

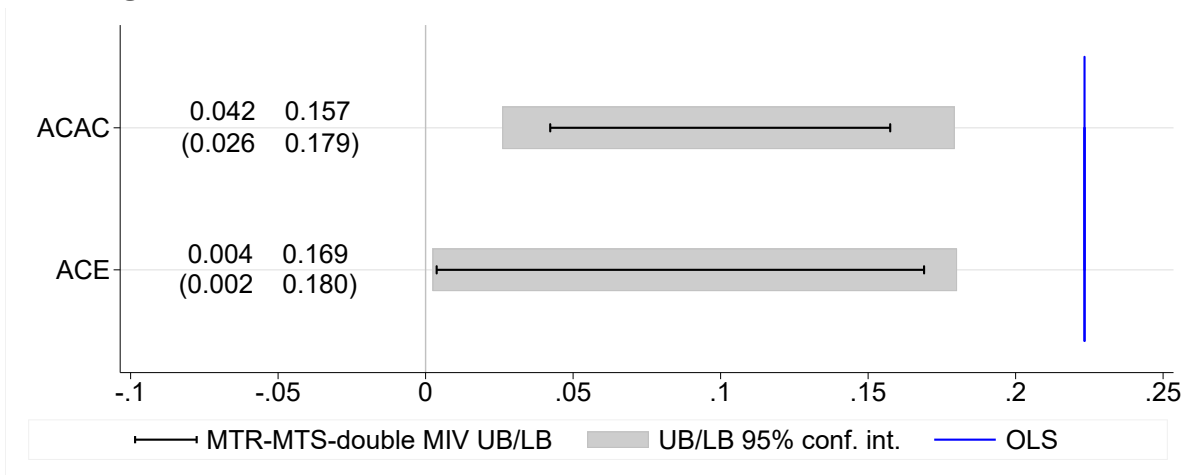
²⁵Comparing Equations 7 and 13 shows that the MTR-MTS upper bound on $E[W^c(1)]$ is identical to the MTS upper bound on $E[W^c(1)]$, and that the MTR-MTS lower bound on $E[W^c(0)]$ is identical to the MTS lower bound on $E[W^c(0)]$. Since the upper bound on the ACE is obtained by subtracting the lower bound on $E[W^c(0)]$ from the upper bound on $E[W^c(1)]$, this implies that the MTR assumption does not affect the upper bound on the ACE. Similarly, as shown in Equation 18, the upper bound on the ACAC depends on the lower bound on $E[W^c(0)]$, so also the upper bound on the ACAC remains unchanged when we impose the MTR assumption.

Figure 10. MTR-MTS-double MIV Bounds Around the ACE and ACAC – DI



Note: Baseline sample of 258,452 children born between 1980 and 1984. The estimated bounds are corrected for finite-sample bias using the method developed by Kreider and Pepper (2007). Confidence intervals are based on the method described in Imbens and Manski (2004) and are obtained from 999 bootstrap replications. To account for correlations in welfare participation between siblings, each replication draws a cluster of siblings. There are 214,401 clusters in the sample.

Figure 11. MTR-MTS-double MIV Bounds Around the ACE and ACAC – FA



Note: Baseline sample of 258 452 children born between 1980 and 1984. The estimated bounds are corrected for finite-sample bias using the method developed by Kreider and Pepper (2007). Confidence intervals are based on the method described in Imbens and Manski (2004) and are obtained from 999 bootstrap replications. To account for correlations in welfare participation between siblings, each replication draws a cluster of siblings. There are 214,401 clusters in the sample.

5 Comparison to Previous Literature

As shown in the previous section, the upper bounds on the ACE and ACAC, for both welfare programs, are well below the corresponding OLS estimates. In this section we further show that the upper bounds are also substantially lower than a number of point estimates from the recent literature. Table 4 gives a schematic overview of the previous literature that aims to identify the causal effect of parents' welfare reciprocity on children's welfare reciprocity. Below we give a brief discussion of the results of each of these papers, and we discuss whether the results are comparable to our findings. Some of the papers are not directly comparable to ours, as they estimate the causal effect of the duration of parental welfare

participation, as opposed to the incidence of participation, while others estimate the effect of a change in the benefit amount. We start by discussing the literature on disability insurance, followed by a discussion of the literature that focuses on welfare programs that are comparable to financial assistance.

5.1 Disability Insurance

The literature on the intergenerational transmission of DI dependency is small and recent. In fact, we are only aware of three studies (listed in the first three rows of Table 4): two using Norwegian data and one using Dutch data.

[Bratberg et al. \(2014\)](#) use Norwegian register data to estimate family fixed effects models, and find that the younger the children are when their parents start receiving DI benefits, the higher the risk that they themselves receive DI benefits at age 40. Due to the nature of the family fixed effect models, all children are exposed to parental DI participation at some point in time. [Bratberg et al. \(2014\)](#) hence estimate a combination of the effect of the duration of receipt, and the effect of the age at exposure to parental DI receipt. This implies that their estimates are not directly comparable to our estimated bounds around the causal effect of parental DI participation on children's DI participation.

The second study on Norwegian data is [Dahl et al. \(2014\)](#) who look at the intergenerational transmission of DI participation in Norway for children whose parents make a court appeal after initially having been denied DI benefits. They estimate a local average treatment effect (LATE) for children whose parent's appeal case would have had a different outcome, had the parent been assigned a more strict or lenient judge. They find that when a parent is granted DI benefits, the probability that the offspring ever receives DI benefits increases by 12 percentage points for children who were aged 18 or above at the time of the decision. For children who were between 8 and 17 years old at the time of the court decision, the probability of applying for DI benefits ten years after their parent's appeal decision increases by 9.5 percentage points if the parent is granted DI benefits.

Since we measure parental DI participation during adolescence, the latter of the two above mentioned estimates is most comparable to our main results. The top panel of Figure 12 shows the point estimate from [Dahl et al. \(2014\)](#), as well as our tightest bounds around the average causal effect of parental DI participation, both for the general population and for the children that grew up in a family with at least one parent on DI. The point estimate from [Dahl et al. \(2014\)](#) is, strikingly, about four times higher than our estimated upper bounds on the ACE and the ACAC.

The third paper on intergenerational spillovers in disability insurance, [Dahl and Gielen \(2018\)](#), exploits a reform in the Netherlands in 1993 that tightened eligibility criteria and lowered payment generosity for DI benefits. [Dahl and Gielen \(2018\)](#) find that the children of parents who, because of the reform, experience a reduction in benefits, or leave the DI program completely are 1.1 percentage points less likely to participate in DI themselves. Since the reform could affect both the intensive and extensive margin of DI reciprocity, it is difficult to directly compare the results from [Dahl and Gielen \(2018\)](#) to our estimates. An approximate, though imperfect, comparison can be made by using their IV estimate of the effect of parental DI benefits on the probability that the child ever participates in DI to compute the effect of decreasing parental DI benefits from the mean of 10,063 euros to zero.²⁶ This gives an approximate

²⁶Another reason the calculation is not perfect is that a drop in benefits from 10,063 euros to zero is very large compared to

Table 4. Previous Literature on the Causal Effect of Parental Welfare Reciprocity on Children’s Welfare Reciprocity

Paper	Country	Welfare program	Method	Treatment			Outcome
				Measure of welfare receipt	Age child	Measure of welfare receipt	
(1) Dahl and Gielen (2018)	Netherlands	DI	RD	Amount	On average 18	Incidence, duration and amount.	On average 36
(2) Dahl et al. (2014)	Norway	DI	2SLS	Incidence	≥ 18 and 7-18	Incidence	≥ 23 and 17-28
(3) Bratberg et al. (2014)	Norway	DI	Family fixed effects	Duration/age of exposure	$\leq 15, 16-20, 21-30, 31-40$	Incidence	40
(4) Hartley et al. (2017)	U.S.	AFDC	2SLS	Incidence	12-18	Incidence	On average 28
(5) Beaulieu et al. (2005)	Canada	Social assistance	Bivariate probit	Duration	7-17	Duration	18-21
(6) Pepper (2000)	U.S.	AFDC	Partial identification	Incidence and duration	12-16	Incidence and duration	19-32
(7) Levine et al. (1996)	U.S.	AFDC	Bivariate probit and IV	Incidence	14-22	Incidence	24-32
(8) Gottschalk (1996)	U.S.	AFDC	Event history	Duration	8-22	Incidence	14-22
(9) Antel (1992)	U.S.	AFDC	Bivariate probit and IV	Incidence	13-18	Duration and incidence	20-27

estimate of the causal effect of parent's DI participation on children's DI participation of 0.09, which is almost identical to the point estimate from [Dahl et al. \(2014\)](#).

Both the estimates from [Dahl et al. \(2014\)](#) and [Dahl and Gielen \(2018\)](#) are substantially larger than our estimated upper bounds. Although it is not possible state with certainty what explains these enormous differences, one likely explanation is that the compliers in both papers differ from the average DI participants in such a way that the average causal effect for children of the complier parents is substantially higher than the average causal effect for the children of average DI participants.

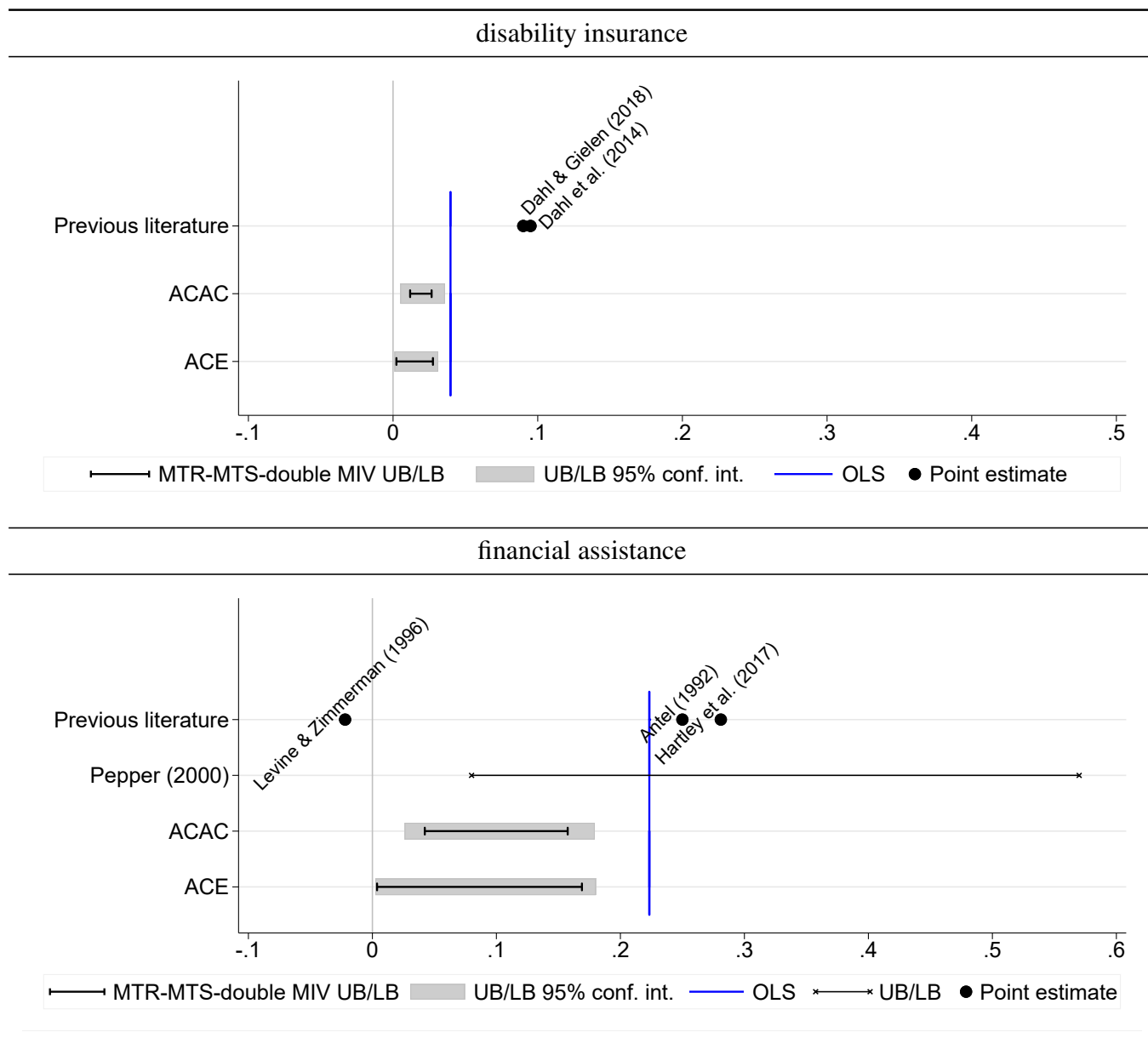
The complier parents in [Dahl et al. \(2014\)](#) are the parents whose DI application would be accepted after their appeal in the case that the appeal is assigned to a lenient judge, but whose DI application would be rejected in case the appeal is assigned to a strict judge. The complier parents in [Dahl and Gielen \(2018\)](#) are the parents who, at age 45, are forced off DI, or whose benefits are reduced if they were to be exposed to the new DI rules, but whose DI benefits remain unchanged if they were not to be exposed to the new DI rules. Both groups of complier parents can be characterized as marginal (potential) DI participants.

The mean characteristics of compliers reported in [Dahl et al. \(2014\)](#) and [Dahl and Gielen \(2018\)](#) show that the marginal DI participants indeed differ on a number of characteristics from the average DI participants. The complier parents in [Dahl and Gielen \(2018\)](#) are for example almost 10% less likely to be fully disabled compared to the average 45-year old DI participant. Around 57% of the 45-year old DI participants in the Netherlands are fully disabled, which is very similar to the share of fully disabled parents in our sample (59%). In contrast, only 48% of the complier parents in [Dahl and Gielen \(2018\)](#) are fully disabled, and in addition they have been on DI for about ten months longer than the average 45-year old DI participant in the Netherlands.

Also the complier parents in [Dahl et al. \(2014\)](#) differ from the average DI participants. Complier parents are for example considerably more likely to have a musculoskeletal or mental diagnosis (termed difficult to verify disorder in [Dahl et al. \(2014\)](#)) than the average DI participants: around 68% of the complier parents have a difficult to verify disorder, while [Dahl et al. \(2014\)](#) report that only 59% of all DI recipients have a difficult to verify disorder. Another notable difference between the complier parents and the average DI participants is the level of education. About 65% of the complier parents are low educated (less than ten years of education), while [Table 2](#) shows that the share of low educated parents among DI participants in general is much lower (53%). In terms of educational attainment, the marginal DI applicants in [Dahl et al. \(2014\)](#) seem more similar to the FA participants, among which 71% are low educated. In that respect it is also interesting to see that while their point estimate is about four times as large as the estimated upper bounds on the ACE and ACAC for DI, it falls well within the bounds around the ACE and ACAC for FA. Heterogeneity in intergenerational spillovers, in combination with marginal DI participants being very different from average DI participants, therefore seems to be a likely explanation for the differences between the point estimates found in the previous literature, and our estimated upper bounds on the ACE and ACAC.

actual drop in benefits due to the reform, which was on average equal to 1,000 Euros.

Figure 12. Comparison of Estimated Bounds to Results from Previous Literature



5.2 Financial Assistance

Also the literature on the causal effect of parental participation in a traditional means-tested welfare program is small, but in contrast to the literature on DI, it is less recent. Most of the existing studies investigate the effect of maternal welfare participation in the Aid to Families with Dependent Children (AFDC) program on daughters' AFDC participation. An important difference between the AFDC program in the U.S. and financial assistance in Norway is that only families with dependent children below the age of 18 are eligible for AFDC, while the presence of children below 18 in the household is not an eligibility criteria for FA. In addition, since the majority of the AFDC recipients are single mothers, the studies investigating the intergenerational transmission of AFDC dependency focus only on mothers and their daughters.

We are aware of six studies that estimate the causal intergenerational transmission of means-tested welfare dependency, which are listed in the bottom six rows of Table 4. The oldest of these studies,

[Antel \(1992\)](#), uses a variant of Heckman’s selection binary endogenous variable estimator. Assuming that characteristics of the state’s welfare system and local labor demand variables (measured when the daughter was between 13 and 18 years old) can be excluded from the daughter’s welfare function, he finds that the mother’s AFDC participation significantly increases her daughter’s expected time on AFDC as a young adult. The result that is most comparable to our estimated bounds is the estimate from the bivariate probit model, evaluated at the sample means, which indicates that having a mother participating in AFDC increases the probability of the daughter’s welfare participation by 25 percentage points.²⁷

[Levine et al. \(1996\)](#) use a similar approach to [Antel \(1992\)](#), but conclude that most – if not all – of the observed correlation between mothers’ and daughters’ receipt can be attributed to the intergenerational correlation in income and other family background characteristics. When assuming that characteristics of the state welfare system and of the local labor market in 1979 do not directly affect daughters’ welfare participation in 1989, Levine et al. find a small negative but statistically insignificant estimate of mothers’ AFDC participation on daughters’ participation.²⁸

In contrast to the previous two papers, [Gottschalk \(1996\)](#) does not rely on exclusion restrictions, but uses instead event history models in combination with assumptions on the timing of events. He starts out by assuming that the mother’s AFDC receipt does not affect the welfare participation of her daughter in earlier periods, and that daughter’s welfare participation does not affect the mother’s welfare participation. In a second empirical approach, [Gottschalk \(1996\)](#) compares the mother’s predicted and actual histories of welfare uptake to draw inferences about unobserved characteristics affecting welfare participation, and he further assumes that these unobservables are correlated across generations. He concludes that unobserved heterogeneity accounts for part of the observed correlation, but that an increase in the mother’s duration of AFDC participation has a significant impact on daughter’s future participation. Since [Gottschalk \(1996\)](#) estimates the effect of the duration of the mother’s welfare participation, the estimates reported in the paper are not directly comparable to our estimated bounds, and are therefore not reported in [Figure 12](#).

[Beaulieu et al. \(2005\)](#) use Canadian administrative data to estimate the intergenerational transmission of reliance on Social Assistance (SA) in Canada. They apply a structural approach in combination with timing of event assumptions similar to those made by [Gottschalk \(1996\)](#), and find a significant causal link between parental and child reliance on SA. Given that their estimated causal link takes into account the impact of both the occurrence and the intensity of parental participation, the findings of this paper are also not directly comparable to our results.

[Pepper \(2000\)](#) is another study estimating the intergenerational transmission of AFDC dependency. As in this paper, [Pepper \(2000\)](#) does not identify a point estimate, but instead exploits nonparametric assumptions to obtain bounds around the average causal effect (ACE) of the mother’s AFDC participation on daughter’s AFDC participation as a young adult. The empirical analysis is based on a sample of 310 black respondents of the PSID that grew up in single parent households. [Pepper \(2000\)](#) shows

²⁷In Section III.B in [Antel \(1992\)](#), he writes the following: “The probit model estimates indicate that the probability of daughter welfare participation in 1985-1987 was increased from 0.07 to 0.32 as AFDCMOM switched from zero to one. (The actual daughter probit AFDCMOM coefficient is 0.8835 with a t-statistic of 2.25.)”. This implies that having a mother participating in AFDC is estimated to increase the probability of daughters’ welfare participation by about 25 percentage points.

²⁸The estimate shown in [Figure 12](#) is the estimate reported in column 4, bottom row of Table 4 in [Levine et al. \(1996\)](#).

that the data alone are not conclusive about the average causal effect of having a mother that received AFDC for a certain duration. He further illustrates what can be learned by combining the data with assumptions motivated by economic theory or empirical convention. The tightest bounds are obtained when he combines an ordered outcome (MTR) assumption with using the local unemployment rate as an instrumental variable.²⁹ Since [Pepper \(2000\)](#) estimates spillover effects for different AFDC spell durations, and he also reports estimates separately for families of different sizes and with different income levels it is difficult to directly compare his results with our estimates. The specification that is most comparable to ours is when he estimates bounds around the average causal effect of growing up in a household that received AFDC for three or four years, as opposed to zero years, on the probability that the daughter participates in AFDC. As we report in [Figure 12](#), he finds an upper and lower bound of 0.08 and 0.57 respectively.³⁰

Finally, a recent study from the U.S. looks at how a welfare reform affected the intergenerational transmission of mothers' AFDC/TNAF participation on daughters' participation. [Hartley et al. \(2017\)](#) instrument mothers' welfare participation by the state maximum AFDC/TNAF benefit levels and the maximum Earned Income Tax Credit measured when daughters were 12 to 18 years old. They find that prior to the welfare reform, the mother's welfare participation increased the likelihood of daughters' participation by around 28 percentage points, but that this transmission was reduced by at least 50 percent due to the reform. Hartley et al. also point out that there is evidence of former AFDC participants enrolling in other welfare programs. When they extend the definition of welfare participation to include these other types of social assistance, they find a pre-reform effect of similar magnitude, but they find no evidence that this transmission channel was changed due to the reform. In [Figure 12](#) we therefore show their pre-reform estimate of the causal transmission of welfare participation.³¹

The bottom panel of [Figure 12](#) shows that a number of point estimates from the previous literature on the intergenerational transmission of means-tested welfare dependency are larger than our estimated upper bounds on both the ACE and the ACAC. As for DI, a likely explanation for the different findings is heterogeneity in spillover effects to children of different sub-populations of recipients. The observed differences might however also be explained by heterogeneity in spillover effects across welfare programs, since AFDC differs from financial assistance both in terms of program characteristics and type of participants.³²

6 Conclusion

The literature on the causal intergenerational transmission of welfare dependency is small, but most of the available papers find results that indicate substantial intergenerational spillovers of welfare programs. These findings are worrying especially considering that many countries have seen dramatic rises in

²⁹Our identification strategy is closely related to that of [Pepper \(2000\)](#), but our paper improves upon the study by [Pepper](#) by using a much larger data set, by focusing not only on the ACE but also the ACAC, and by imposing weaker assumptions, for example by using local labor market conditions as a monotone instrumental variable instead of an instrumental variable, and thereby relaxing the mean independence assumption to a mean monotonicity assumption.

³⁰The estimated bounds are reported in row VI.C in Table 9 of [Pepper \(2000\)](#).

³¹The point estimate reported in [Figure 12](#) is the estimate reported in row 1, column 2 in Table 2 of [Hartley et al. \(2017\)](#)

³²Another reason for differences in findings could be the use of different methods based on different sets of identifying assumptions.

the number of individuals receiving social assistance or disability benefits: On average in the OECD countries, the number of DI recipients increased by about 8 percent from 2007 to 2014, and the number of individuals receiving benefits under a social assistance program increased by about 14 percent during the same time period.³³ Substantial intergenerational spillovers might be one explanation for this upward trend in welfare participation, and this might also imply that we will see a continued or even greater increase in the number of welfare recipients in the future (as the children of today's participants start claiming benefits). This reasoning is, however, based on the idea that the estimates of intergenerational spillovers for particular sub-groups of welfare participants reported in the literature can be extrapolated to the broad population of welfare recipients.

The results in this paper show that the average causal effect of parents' welfare participation on children's welfare participation as an adult, both in the full population and in the population of children that grew up in welfare-dependent families, is considerably smaller than many of the estimates reported in the previous literature. These findings indicate that there is important heterogeneity in the intergenerational spillovers of welfare programs and that these spillovers are much larger for children of marginal welfare participants compared to children of average welfare participants. We further find that a substantial share of the observed correlation between parents' and children's welfare participation is due to shared characteristics. When imposing the MTR assumption we do however find positive lower bounds that are significantly different from zero. These bounds also indicate that the intergenerational spillovers of the means-tested FA program are larger than the intergenerational spillovers from DI.

Spillovers of welfare programs onto the offspring of average welfare participants might warrant policy measures directly targeting the children of welfare recipients. However, in order to know what type of policy measures might be effective, we need more research into the mechanisms behind the observed intergenerational spillovers. These mechanisms are likely to be related to parents' responses to receiving (as opposed to not receiving) welfare benefits, i.e. the parents' counterfactual outcomes. Imagine for example, that parents would work in the case that they do not receive welfare benefits, while they would leave the labor market in the case that they are granted a benefit. In this scenario, a possible mechanism behind the causal effect of parents' welfare participation on children's welfare participation is parents' inability to transfer job search skills and informal job contacts to their offspring. Moreover, as marginal welfare participants might be more capable of working than average welfare participants, the counterfactual outcomes of marginal welfare participants might differ from the counterfactual outcomes of average welfare participants. More information about parents' counterfactual outcomes, both for the marginal participants and for the full population of welfare participants, is hence necessary both to give insights into the mechanisms behind the observed heterogeneity in the intergenerational spillovers of welfare programs, and to design potential policy measures targeting the children of welfare participants.

³³Source: OECD SOCR database. <http://www.oecd.org/social/recipients.htm>

References

- Antel, J. J. (1992). The intergenerational transfer of welfare dependency: Some statistical evidence. *The Review of Economics and Statistics*, 74(3):467–473.
- Beaulieu, N., Duclos, J.-Y., Fortin, B., and Rouleau, M. (2005). Intergenerational reliance on social assistance: Evidence from Canada. *Journal of Population Economics*, 18(3):539–562.
- Björklund, A. and Salvanes, K. G. (2011). Education and family background: Mechanisms and policies. In *Handbook of the Economics of Education*, volume 3, pages 201–247. Elsevier.
- Black, S. and Devereux, P. (2011). Recent Developments in Intergenerational Mobility. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 4B, chapter 16, pages 1487–1541. Elsevier, 1 edition.
- Bragstad, T. (2009). Tidsbegrenset uførestønad evaluering fire år etter innføring. NAV-rapport 3/2009.
- Bratberg, E., Nilsen, i. A., and Vaage, K. (2014). Assessing the intergenerational correlation in disability pension recipiency. *Oxford Economic Papers*, 67(2):205–226.
- Currie, J. (2009). Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. *Journal of Economic Literature*, 47(1):87–122.
- Dahl, G. B. and Gielen, A. C. (2018). Intergenerational spillovers in disability insurance. Technical report, National Bureau of Economic Research.
- Dahl, G. B., Kostøl, A. R., and Mogstad, M. (2014). Family Welfare Cultures. *The Quarterly Journal of Economics*, 129(4):1711–1752.
- Ellingsen, J. (2017). Utviklingen i uføretrygd per 30. september 2017. NAV.
- Fiva, J. H., Halse, A., and Natvik, G. J. (2017). Local government dataset. *Dataset available at www.jon.fiva.nl/data.htm*.
- Gottschalk, P. (1996). Is the correlation in welfare participation across generations spurious? *Journal of Public Economics*, 63(1):1–25.
- Hartley, R. P., Lamarche, C., and Ziliak, J. P. (2017). Welfare reform and the intergenerational transmission of dependence. Technical report, IZA Discussion Paper No. 10942.
- Imbens, G. W. and Manski, C. F. (2004). Confidence intervals for partially identified parameters. *Econometrica*, 72(6):1845–1857.
- Immerwoll, H. (2012). *Minimum-Income Benefits in OECD Countries: Policy Design, Effectiveness and Challenges*. In: *Counting the Poor. New Thinking About European Poverty Measures and Lessons for the United States*. Oxford University Press.
- Kreider, B. and Pepper, J. V. (2007). Disability and employment: reevaluating the evidence in light of reporting errors. *Journal of the American Statistical Association*, 102(478):432–441.
- Levine, P. B., Zimmerman, D. J., et al. (1996). *The Intergenerational Correlation in AFDC Participation: Welfare Trape Or Poverty Trap?* Institute for Research on Poverty, University of Wisconsin-Madison.

- Manski, C. F. (1989). Anatomy of the selection problem. *Journal of Human Resources*, 24(3):343–360.
- Manski, C. F. and Pepper, J. V. (2000). Monotone instrumental variables: with an application to the returns to schooling. *Econometrica*, 68(4):997–1010.
- Manski, C. F. and Pepper, J. V. (2009). More on monotone instrumental variables. *The Econometrics Journal*, 12(s1):S200–S216.
- Markussen, S. and Røed, K. (2016). Leaving poverty behind? the effects of generous income support paired with activation. *American Economic Journal: Economic Policy*, 8(1):180–211.
- Mogstad, M., Santos, A., and Torgovitsky, A. (2017). Using instrumental variables for inference about policy relevant treatment effects. Technical report, National Bureau of Economic Research.
- NAV (2012). Rundskriv hovednummer 35 sosialtjenesteloven, the national insurance act, accessed 01.12.2017.
- NAV (2014). Rapportering om kvalifiseringsprogrammet 12.02.2015, accessed 19.12.17.
- OECD (2010). *Sickness, Disability and Work: Breaking the Barriers. A Synthesis of Findings across OECD Countries*. OECD, Paris.
- Page, M. E. (2004). *New evidence on the inter generational correlation in welfare participation*. Cambridge University Press.
- Pepper, J. V. (2000). The intergenerational transmission of welfare receipt: A nonparametric bounds analysis. *Review of Economics and Statistics*, 82(3):472–488.
- Rege, M., Telle, K., and Votruba, M. (2009). The effect of plant downsizing on disability pension utilization. *Journal of the European Economic Association*, 7(4):754–785.

A Appendix For Online Publication

Table A1. Main Programs in the Norwegian National Welfare System

Program	Norwegian program name
<i>Disability insurance</i>	Uføretrygd Foreløpig uføretrygd
<i>Financial assistance</i>	Økonomisk sosialhjelp
<i>Unemployment benefits</i>	Dagpenger
<i>Old-age pension</i>	Alderspensjon
<i>Sickness benefits</i>	Sykepenger
<i>Vocational rehabilitation programs</i>	Attføring og rehabiliteringspenger
	Tidsbegrenset uførestønad
	Arbeidsavklaringspenger (merger of the two above mentioned programs)
<i>Qualification program</i>	Kvalifiseringsprogrammet

Figure A1. Construction of the Unemployment MIV with Varying numbers of Categories

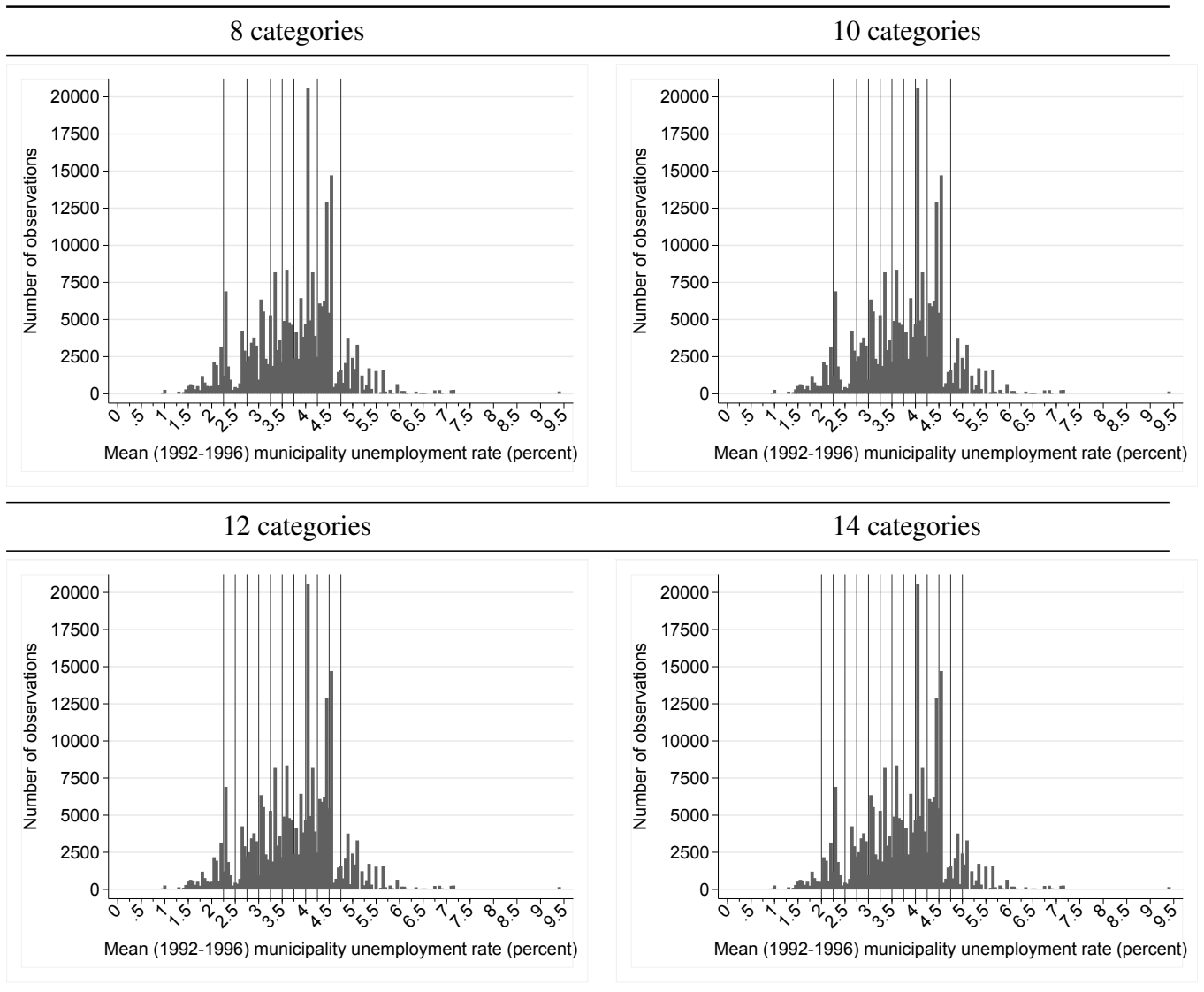


Table A2. Parental Characteristics by Welfare Dependency within Subsamples Defined by the MIV's

MIV: unemp. rate:		MIV: parent's level of education											
		Compulsory				Completed upper secondary				Completed tertiary education			
		DI	$w^p = 1$	$w^p = 0$	FA	DI	$w^p = 1$	$w^p = 0$	FA	DI	$w^p = 1$	$w^p = 0$	FA
<2.25%	Foreign born	0.05	0.05	0.29	0.05	0.13	0.08	0.48	0.07	0.19	0.10	0.83	0.10
	Married	0.69	0.86	0.47	0.86	0.77	0.88	0.62	0.88	0.82	0.91	0.80	0.90
	Earnings	1.33	4.40	0.94	4.28	1.69	5.56	0.88	5.51	2.23	6.88	0.45	6.84
2.25%-2.5%	Foreign born	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$
	Married	0.12	0.08	0.28	0.08	0.13	0.12	0.47	0.12	0.25	0.15	0.67	0.15
	Earnings	0.62	0.82	0.29	0.82	0.65	0.84	0.47	0.84	0.72	0.88	0.52	0.88
2.5%-2.75%	Foreign born	1.33	4.41	1.06	4.27	1.69	5.97	1.48	5.92	2.23	7.72	0.8	7.66
	Married	0.06	0.05	0.22	0.05	0.1	0.07	0.55	0.07	0.13	0.11	0.75	0.1
	Earnings	0.72	0.85	0.39	0.85	0.81	0.87	0.55	0.87	0.84	0.92	0.63	0.92
2.75%-3%	Foreign born	1.21	4.23	0.72	4.07	1.95	5.6	0.47	5.53	2.33	6.89	1.15	6.83
	Married	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$
	Earnings	0.07	0.06	0.22	0.06	0.11	0.07	0.64	0.07	0.19	0.11	0.57	0.11
3%-3.25%	Foreign born	0.72	0.83	0.41	0.83	0.66	0.84	0.64	0.84	0.79	0.89	0.48	0.89
	Married	1.2	4.19	0.79	4.03	2.02	5.8	0.7	5.72	2.02	6.64	0.93	6.58
	Earnings	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$
3.25%-3.5%	Foreign born	0.07	0.06	0.19	0.05	0.09	0.08	0.43	0.07	0.2	0.09	0.56	0.09
	Married	0.63	0.77	0.29	0.77	0.68	0.81	0.45	0.81	0.77	0.84	0.41	0.84
	Earnings	1.1	4.16	0.57	3.99	1.82	5.62	0.76	5.56	2.21	6.8	0.7	6.74
3.5%-3.75%	Foreign born	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$
	Married	0.08	0.06	0.18	0.06	0.09	0.08	0.45	0.08	0.18	0.1	0.78	0.1
	Earnings	0.66	0.8	0.32	0.8	0.74	0.84	0.6	0.84	0.75	0.89	0.75	0.89
3.75%-4%	Foreign born	1.06	4.12	0.72	3.97	1.73	5.63	0.54	5.57	2.46	6.91	0.64	6.86
	Married	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$
	Earnings	0.08	0.06	0.18	0.06	0.09	0.08	0.45	0.08	0.18	0.1	0.78	0.1

MIV: parent's level of education

MIV: unemp. rate:	Compulsory						Completed upper secondary						Completed tertiary education						
	DI	$w^p = 1$	$w^p = 0$	FA	$w^p = 1$	$w^p = 0$	DI	$w^p = 1$	$w^p = 0$	FA	$w^p = 1$	$w^p = 0$	DI	$w^p = 1$	$w^p = 0$	FA	$w^p = 1$	$w^p = 0$	
3.5%-3.75%	Foreign born	0.09	0.07	0.28	0.07	0.07	0.16	0.09	0.47	0.08	0.22	0.12	0.72	0.11					
	Married	0.63	0.79	0.35	0.79	0.83	0.66	0.83	0.41	0.83	0.74	0.88	0.54	0.88					
	Earnings	1.05	4.06	0.71	3.93	5.61	1.92	5.61	0.69	5.57	2.5	7.02	0.88	6.97					
3.75%-4%	Foreign born	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 0$
	Foreign born	0.07	0.07	0.24	0.06	0.06	0.14	0.08	0.51	0.08	0.16	0.1	0.67	0.1					
	Married	0.63	0.77	0.34	0.77	0.81	0.69	0.81	0.42	0.81	0.74	0.86	0.61	0.86					
Earnings	1.02	3.85	0.62	3.66	5.11	1.94	5.2	0.68	5.11	2.05	6.44	0.58	6.39						
4%-4.25%	Foreign born	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 0$
	Foreign born	0.29	0.18	0.44	0.18	0.18	0.37	0.17	0.62	0.17	0.43	0.18	0.82	0.18					
	Married	0.62	0.73	0.37	0.74	0.77	0.67	0.77	0.45	0.78	0.67	0.81	0.61	0.81					
Earnings	0.81	3.77	0.6	3.62	5.48	1.5	5.54	0.82	5.48	1.49	7.33	0.64	7.25						
4.25%-4.5%	Foreign born	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 0$
	Foreign born	0.08	0.07	0.23	0.06	0.06	0.13	0.08	0.42	0.07	0.14	0.1	0.71	0.1					
	Married	0.61	0.75	0.3	0.75	0.8	0.63	0.8	0.39	0.8	0.72	0.85	0.56	0.85					
Earnings	0.83	3.83	0.54	3.68	5.27	1.47	5.33	0.62	5.27	2.21	6.91	0.78	6.86						
4.5%-4.75%	Foreign born	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 0$
	Foreign born	0.11	0.09	0.24	0.08	0.08	0.2	0.1	0.45	0.1	0.14	0.11	0.63	0.11					
	Married	0.59	0.73	0.27	0.74	0.78	0.72	0.78	0.44	0.79	0.65	0.85	0.5	0.85					
Earnings	0.89	3.91	0.58	3.79	5.35	1.26	5.42	0.76	5.35	1.68	7.26	0.83	7.21						
>4.75%	Foreign born	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 1$	$w^p = 0$	$w^p = 0$
	Foreign born	0.06	0.05	0.14	0.05	0.05	0.1	0.06	0.42	0.06	0.12	0.09	0.69	0.08					
	Married	0.61	0.74	0.25	0.74	0.78	0.72	0.78	0.4	0.79	0.76	0.86	0.61	0.86					
Earnings	0.89	3.76	0.59	3.6	5.11	1.75	5.16	0.7	5.11	2.17	6.54	0.9	6.5						