CERGE-EI

Center for Economic Research and Graduate Education -
Economics Institute
A joint workplace of Charles University and the Economics Institute
Czech Academy of Sciences

Master thesis

2022 **Bc. Sofiana Sinani**

CERGE-EI

Center for Economic Research and Graduate Education -
Economics Institute
A joint workplace of Charles University and the Economics Institute
Czech Academy of Sciences

Bc. Sofiana Sinani

The Labor Market Effects of the Working Family Tax Credit on Single Mothers in the UK

Master thesis

Prague 2022

Author: Bc. Sofiana Sinani

Supervisor: Andreas Menzel, Ph.D.

Academic year: 2020/2021

Bibliographic note

SINANI, Sofiana. The Labor Market Effects of the Working Family Tax Credit on Single Mothers in the UK. Prague 2022. Master thesis. Center for Economic Research and Graduate Education – Economics Institute, A joint workplace of Charles University and the Economics Institute of the Czech Academy of Sciences.
Thesis supervisor: Andreas Menzel, Ph.D.

Abstract

The Working Family Tax Credit (WFTC) was introduced in the UK in 1999. This policy provided financial incentives for work to low-income parents with children (primarily single mothers). The parents received full credit if they worked 16 hours/week and the credit decreases with more hours of work. The credit increased marginally when the parent worked 30 hours/week to stimulate full-time work. The WFTC unambiguously made work attractive; however, the effect on labor hours for parents employed in pre-WFTC is unclear. In this thesis, I investigate the effect of the WFTC on single mothers' paid labor hours. I apply the quantile difference-in-difference approach and estimate the quantile treatment effects. This approach allows me to find the effect of the WFTC given the single mothers' paid labor hours in pre-WFTC. My results highlight the existence of treatment heterogeneity. The QTE suggests that the effect of the WFTC was significantly positive for single mothers in the bottom quantiles, while for the upper quantiles there was either no effect (unconditional QTE) or a negative effect (conditional QTE). My findings are consistent with a labor supply model where single mothers in the bottom quantiles work more because of the substitution effect, while those in the upper quantiles substitute labor for leisure because of the income effect. The income effect could potentially be explained by the withdrawal rate and the interaction of the WFTC with other social benefits. Overall, my findings suggest that the WFTC shifted more single mothers into part-time employment (16 hours/week), but not into full-time employment (30 hours/week). Nevertheless, I cannot confidently say that my effects do not pick up the effect of other policies that were introduced shortly before the WFTC.

Key words: labor hours, WFTC, single mothers, quantile, income effect, substitution effect.

Abstrakt

Working Family Tax Credit (WFTC) byla sociální politika zavedená ve Spojeném Království v roce 1999. Tato politika vytvořila systém finančních pobídek pro rodiče s nízkými příjmy (hlavně matky samoživitelky). Pokud rodiče odpracovali alespoň 16 hodin za týden, obdrželi plný kredit. S rostoucím počtem odpracovaných hodin velikost získaného kreditu klesala, avšak skokově vrostla při odpracování 30 hodin za týden a následně znovu stejným tempem klesala. WFTC přiměla pracovat pouze ty rodiče, kteří před zavedením WFTC nepracovali. Vliv WTFC na počet odpracovaných hodin rodičů pracujících i před zavedením politiky je nejasný. V práci se zabývám vlivem WFTC na počet odpracovaných hodin matek samoživitelek. Využívám k tomu kvantilovou metodu rozdílu v rozdílech (dále jen QTE), pomocí níž odhaduji efekt politiky pro jednotlivé kvantily. Metoda umožňuje odhalit vliv WFTC na počet odpracovaných hodin matek samoživitelek, které pracovaly již před zavedením této politiky. Výsledky práce poukazují na přítomnost heterogenity. QTE naznačuje, že zavedení WFTC mělo pozitivní vliv na matky samoživitelky spadající do spodních kvantilů. Naopak u těch, které spadají do horních kvantilů, nebyl pozorovaný efekt žádný (nepodmíněný QTE), resp. negativní (podmíněný QTE). Tyto závěry jsou konformní s modelem nabídky práce, v němž matky samoživitelky spadající do spodních kvantilů, pracují více, a tak u nich převažuje substituční efekt. Naopak matky samoživitelky spadající do horních kvantilů si kupují více volného času, protože u nich převládá důchodový efekt. Důchodový efekt může být vysvětlen i interakcí WFTC s jinými sociálními benefity. V neposlední řadě, práce naznačuje, že WFTC motivuje matky samoživitelky pracovat na částečný úvazek (16 hodin za týden), ale už ne na plný úvazek (30 hodin týdně). Závěry této práce můžou být zkreslené, protože nebylo možné oddělit efekt WFTC od efektu jiných sociální politik zavedených ve stejném období.

Klíčová slova:: pracovní doba, WFTC, svobodné matky, quantile, příjmový efekt, substituční efekt.

Declaration of Authorship

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree. I hereby proclaim that I wrote my master thesis on my own under the leadership of my supervisor and that the references include all resources and literature I have used.

I grant a permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, August 2, 2022

Signature

Acknowledgments

I would like to thank my supervisor Andreas Menzel, PhD for his continuous support and helpful comments. I am grateful that I had the opportunity to work under his guidance. I would also like to thank Nikolas Mittag for the econometric knowledge that he taught me, and Deborah Novakova for helping me become a better writer. Last but not least, I thank my mother and my father who supported me during my master's degree. I could not have done it without them.

Project of Master Thesis

Author of the master thesis:	**Bc. Sofiana Sinani**
Supervisor of the master thesis:	**Andreas Menzel, Ph.D.**
Academic year:	**2021/2022**

Theme: The Labor Market Effects of the Working Family Tax Credit on Single Mothers in the UK

Research question and motivation:

This thesis aims to explore the labor market effects of the Working Family Tax Credit (WFTC) on the employment of single mothers in the UK. The WFTC was introduced in October 1999 as an anti-poverty strategy and was designed to make labor force participation more attractive. The WFTC gave financial incentives to parents if they worked 16 hours or more per week. A family was considered eligible for the WFTC if they met three requirements: first, at least one parent in the household worked 16 hours or more per week; second, the parent(s) had at least one dependent child; and third, the household's savings were less than £8,000, while savings between £3,000 and £8,000 reduce the credit. Single mothers were the primary recipients of the WFTC (Francesconi and Van der Klaauw, 2007). There is a large literature studying the effects of the WFTC on employment and other labor-market-related variables. The majority of the literature follows a Difference-In-Difference identification strategy (DID), using single women without children as their control group (Brewer et al., 2005; Leigh, 2005; Blundell et al., 2005; Francesconi and van der Klaauw, 2004). Nevertheless, few studies find the effect of the WFTC on paid labor hours. Microeconomic theory would suggest that the WFTC created ambiguous labor supply effects for those who were in employment pre-WFTC. On the one hand, labor income was higher, implying that if the single mother worked more hours she would earn more income. This behavior is called the substitution effect, which implies that the WFTC led to an increase in the number of labor hours. On the other hand, the mother could work less while maintaining the same level of income as in pre-WFTC. This behavior is the income effect, which leads to a decline in the single mothers' paid labor hours. The income and substitution effects, two opposing forces, potentially create heterogeneous treatment effects regarding the hours of work the single mother was initially working in pre-WFTC. The aim of this thesis is to find the effect of the WFTC on the single mothers' paid labor hours.

Contribution:

To the best of my knowledge, no prior study analyzes the heterogeneous treatment effect of the WFTC on the number of labor hours. Gregg and Harkness (2003) measure the mean impact of the WFTC on labor hours for those who entered employment because of the WFTC. They find that the average effect of the WFTC was an increase of 2.5 hours/week. However, the

variance of paid labor hours is high compared to its' mean, implying that many women lie on different ends of the distribution. This implies that the mean impact conceals significant heterogeneity. In this thesis, I address this literature gap by analyzing the heterogeneous treatment effects of the WFTC on the single mothers' paid labor hours. My contribution is two-fold. It is the first study in the WFTC literature that estimates the heterogeneous treatment effects regarding the initial paid labor hours. This offers insight regarding the dynamics of the substitution and income effects. Second, my findings suggest that the WFTC shifted single mothers into part-time employment but not into full-time employment.

Methodology:

I use data from the UK Quarterly Labor Force Survey from Spring 1996 - Winter 2002. My outcome variable is paid labor hours in the week the woman was interviewed in. My sample consists of all single women of working age who were employed in pre-WFTC. I apply a quantile difference-in-difference identification strategy and estimate the quantile treatment effects (Athey and Imbens, 2006). I use single women without children without as my control group because they were not eligible for the policy. A quantile shows where on the inverse cumulative distribution of paid labor hours a mother is located in. Using this approach, I am able to estimate the effect of the WFTC at any given quantile. For example, the 30th quantile in pre-WFTC corresponds to 14 hours/week of paid labor hours.

Outline:

1. Introduction
2. Structure of In-Work Benefits in the US and the UK
3. A Theory of Labor Supply
4. Data
5. The Quantile Difference-In-Difference Approach
6. Results
7. Discussion
8. Conclusion

References:

1. Athey, Susan, and Guido W. Imbens. (2006). "Identification and Inference in Nonlinear Difference-In-Differences Models." Econometrica 74 (2): 431–97. doi.org/10.1111/j.1468-0262.2006.00668.x.

2. Blundell, R., M. Brewer, and A. Shephard. (2005). "Evaluating the Labour Market Impact of Working Families' Tax Credit Using Difference-In-Differences." University College London.

London, UK. June 1, 2005.

3. Brewer, Mike, Alan Duncan, Andrew Shephard, and Maria Jose Suarez. (2005). "Did Working Families' Tax Credit Work? The Final Evaluation of the Impact of In-Work Support on Parents' Labour Supply and Take-up Behaviour in the UK." Institute of Fiscal Studies. June 17, 2005.

4. Francesconi, M, and W Van der Klaauw. (2007). "The Socioeconomic Consequences of 'In-Work' Benefit Reform for British Lone Mothers." Journal of Human Resources. 42,1-31. 2007.

5. Gregg, P, and S Harkness. (2003). "Welfare Reform and Lone Parents Employment in the UK." Centre for Market and Public Organisation. 2003.

6. Leigh, Andrew. (2005). "Optimal Design of Earned Income Tax Credits: Evidence from a British Natural Experiment." Center for Economic Policy Research.

Contents

[1]

[1]Previous versions of this paper were submitted in *Research Writing 1*, *Academic Writing 2*, *Microeconomet-*

1 Introduction

Employment rates in the UK during the 1990s rose dramatically, primarily due to the rapid growth of the employment rates of married mothers (Gregg and Harkness, 2003). However, the employment rates of single mothers were lower in the 1990s than in the 1970s. The UK was one of the few developed nations that faced such low employment rates for single mothers, with fewer than 40 percent of them participating in the labor market (OECD, 1996). These high unemployment rates led to more British children living in poor, jobless households.

During the 1997 political campaign in the UK, the Labor government promised ambitious initiatives to eradicate child poverty, motivated by the general rise of income inequality in the decades that preceded the elections. The poverty standard for UK households includes those households with incomes of less than 60 percent of the national median income, adjusted for household size (Brewer and Gregg, 2003). As of 1998, 33 percent of British children were living in poverty. This number rose by 14 percent relative to 1979. Brewer and Gregg (2003) discuss three primary reasons that led to the increase in child poverty. First, there was a 9 percent increase of jobless households from 1979 to 1998. Simultaneously, the number of households where all adults were working increased. This dichotomy suggests that there was an overall rise in households with a single adult member, implying that the share of single-parent households surged. The second reason could be that there was an increase in income inequality. Among all developed nations during 1979 and 1996, the UK was the only European country to experience the highest increase in the ratio of the 90th percentile of income over the 10th percentile (OECD, 1996). Third, nearly 20 percent of British children lived in jobless households (mainly single-parent), almost double the share of any other European country. This encouraged the UK government to adopt a series of family tax policies to alleviate poverty and encourage employment among those households, particularly single-parent households. With the share of single parents rising over the previous decades, the primary recipients of family tax policies were single parents (Meghir et al., 2009). Single mothers, particularly, were of special interest to policymakers because many tax policies targeted them.

One of the strategies the Labor government introduced to reduce the share of jobless families and child poverty was the Working Family Tax Credit (WFTC) in October 1999. The WFTC gave financial incentives to parents if they worked 16 hours or more per week. A family was considered eligible for the WFTC if they met three requirements: first, at least one parent in the household worked 16 hours or more per week; second, the parent(s) had at least one dependent child[2]; and third, the household's savings were less than £8,000, while savings between £3,000 and £8,000 reduce the credit. The WFTC replaced the Family Credit (FC) in 1999 and became the main form of state support for low-income working families with children until 2003, when it was replaced by the Working Tax Credit. The WFTC covered 70 percent

rics 2 during Spring 2021, Fall 2021, and Spring 2022 at CERGE-EI.

[2]A dependent child was a child, grandchild, stepchild, or foster child under the age of 16 or 19 if they were enrolled in full-time education

more families than the FC.

There is a large literature studying the labor market effects of the WFTC on single mothers. The WFTC was designed to make labor force participation more attractive; hence, the primary outcome of interest has been the employment rate of single mothers. The majority of the literature follows a Difference-In-Difference identification strategy (DID), using single women without children as the main control group. The Average Treatment Effect estimates (ATE) range from 1.4 to 7 percentage points increase in the employment rate after the introduction of the WFTC (Brewer et al., 2005; Leigh, 2005; Blundell et al., 2005; Francesconi and van der Klaauw, 2004). The literature also finds heterogeneity regarding the number of children in an eligible household (Gregg and Harkness, 2003; Blundell et al., 2005) and by the age of the youngest child (Francesconi and Van der Klaauw, 2007; Blundell et al., 2005; Gregg and Harkness, 2003). Gregg and Harkness (2003) also measure the ATE on hours of work, earnings, and poverty. They find that the ATE of the WFTC on hours of work is 2.5 hours/week, which is mainly driven by the new entrants into the labor market. On the other hand, Francesconi and Van der Klaauw (2007) find that the WFTC had no effect on the incumbents' labor hours, while Meyer and Rosenbaum (2001) predicted that the EITC would marginally decrease incumbents' labor hours. Although the WFTC literature agrees that the WFTC increased labor market participation, there is less consensus on its effects on labor hours. Microeconomic theory would suggest that the WFTC created ambiguous labor supply effects for those who were in employment pre-WFTC. On the one hand, the returns to working more hours were higher under the WFTC; thus, more labor hours led to more income. This is the *substitution effect*, which implies that the WFTC leads to an increase in the number of labor hours. On the other hand, because the returns to working were higher, working the same number of hours as in pre-WFTC led to more income. Therefore, the mother may have decreased the number of hours she worked to maintain the previous level of income. This is called the *income effect*, which leads to a decline in the single mothers' paid labor hours. The income and substitution effects, two opposing forces, potentially create heterogeneous treatment effects regarding the hours of work the single mother was initially working in pre-WFTC. Disentangling these effects offers insights to policymakers to enhance the design of future programs that stimulate employment. To the best of my knowledge, no prior study analyzes the heterogeneous treatment effect of the WFTC on the number of labor hours.

In this thesis, I address this literature gap by analyzing the effect of the WFTC on the single mothers' paid labor hours. My identification strategy follows the Quantile Difference-in-Difference approach (QDID) as proposed by Athey and Imbens (2006). I use single women without children as my control group because they were not eligible for the policy. The QDID uses the inverse Cumulative Distribution Function (CDF) of any given variable, which is a set of quantiles that shows the location of an individual on the distribution. For instance, the 30th quantile in the distribution of the single mothers' paid labor hours pre-WFTC corresponds to 14 hours/week. This approach allows me to estimate the effect of the WFTC given the initial

quantile of a single mother, assuming she remains on the same quantile post-WFTC. I estimate the unconditional and conditional[3] quantile treatment effects (QTE) to examine how single mothers reacted to the WFTC at any given quantile. My findings confirm that the mean impact disguises significant heterogeneity. The unconditional and conditional QTE suggest that the WFTC significantly increased the single mothers' paid labor hours for mothers on the bottom quantiles of the paid hours distribution by between 2-6 hours/week[4]. However, the effect on the top quantiles (those working more than 30 hours/week pre-WFTC) varies by specification: the unconditional QTE finds no effect, while the conditional QTE suggests that the WFTC reduced paid labor hours by 2-5 hours/week. The effect on the median is small and varies by specification. My findings are consistent with a labor supply model where the substitution effect encouraged single mothers in the bottom quantiles to increase their labor hours, while the income effect discourages mothers from work because of the WFTC withdrawal rate and because of the interaction of WFTC with other social benefits. Overall, I find that the WFTC was effective into pushing more mothers into part-time employment (16 hours/week) but not full-time employment (30 hours/week).

The contribution of this thesis is twofold. To the best of my knowledge, it is the first study in the WFTC literature to analyze treatment heterogeneity regarding initial paid labor hours. Analyzing this allows me to find clearer evidence of the effect of the WFTC and of the substitution and income effects. Second, my findings suggest that the WFTC encouraged single mothers to work part-time, but did not encourage more single mothers into full-time employment. This could bring new insights to policymakers that aim to design better incentives to increase labor supply in low-income families. I suggest that the WFTC could have increased paid labor hours more if the government concentrated on increasing the WFTC reward at the starting level (16 hours/week), rather than spreading the coverage across the paid labor hours distribution.

The rest of the thesis is organized as follows: Section 2 provides a review of in-work benefits in the US and the UK, and a short history of the family tax credits in the UK; Section 3 introduces a simple model of labor supply, Section 4 describes the data used in the analysis and provides summary statistics; Section 5 introduces the QDID approach and the identification of the QTE estimator; Section 6 introduces the results, 7 discusses the results; and Section 8 concludes.

[3]The conditional QTE is the adjusted specification for other labor-market-related variables.

[4]Depending on the quantile.

2 Structure of In-Work Benefits in the US and the UK

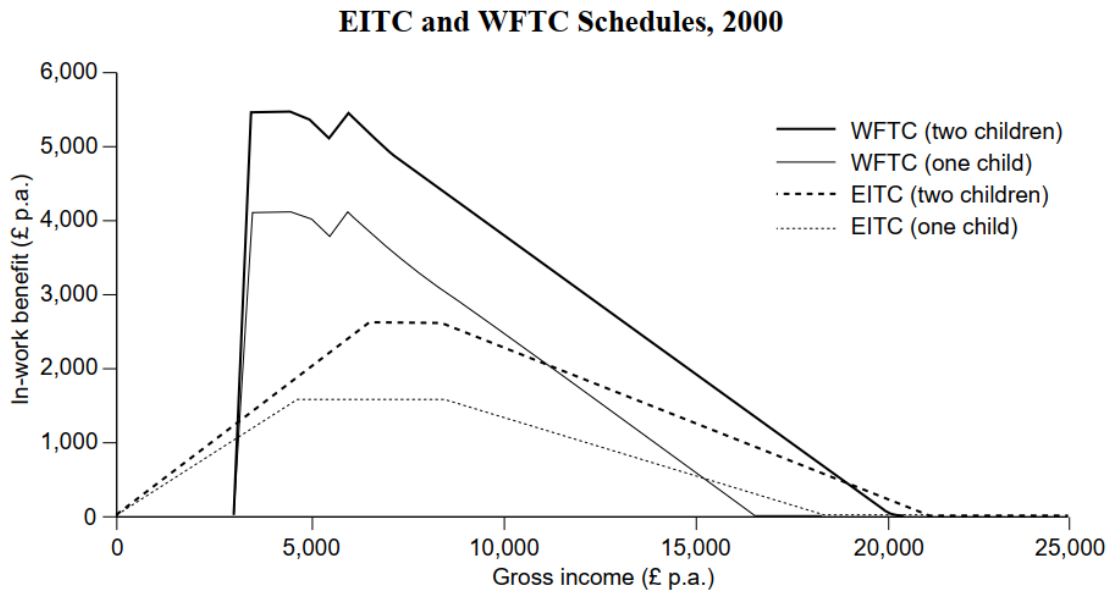
2.1 Evidence from the US Reforms

Anti-poverty programs began in the US with the Aid to Families with Dependent Children (AFDC) in 1935, which transferred cash to families in poverty. The welfare system expanded to a variety of programs, including Food Stamps, Medicaid, and housing benefits, primarily targeting single parents facing limited budgets (Blundell and Hoynes, 2001). These "out-work" welfare programs did not offer any financial incentives to "make work pay" – families received the maximum benefit if they were not participating in the labor market, while benefits were significantly reduced with positive earnings. Hence, their effects on the parents' labor supply decisions were adverse.

The Earned Income Tax Credit (EITC) began in 1975 as an "in-work" benefit with the intent to offset the social security payroll tax for low-income families with children (Hoffman and Seidman, 2003). It differed from previous welfare programs in three aspects: first, the EITC was provided through the tax system (the Internal Revenue Service) rather than the welfare system; second, it was available to all low-income families despite the marital status of the household head; and third, it required that the household head made some positive income during the previous year. Families without earnings would receive nothing, emphasizing its incentive to push parents into employment. In essence, the EITC was the opposite of previous welfare programs which offered maximum benefits to unemployed parents. The EITC eligibility structure was fully reformed by 1996. Eligibility for the EITC depended on the annual earned income and the number of dependent children. The maximum allowance, which was received as a lump-sum tax at the end of the fiscal year, was doubled since 1975 and it exceeded $\$3,000$ for parents with two children (Hoffman and Seidman, 2003), while the take-up rate increased by nearly 70 percent. The amount that each family received depended on the household heads' earnings and the number of dependent children. There are three regions in the credit schedule (figure 1). The initial region was the phase-in region, which transferred cash equal to the subsidy rate times the parents earnings. The second region was the maximum amount a family could receive based on the number of dependent children, ranging from $\$3,000$ to $\$7,000$ per year (Tyger, 2019). The third region was the phase-out region, where the credit was phased out at the same rate for all families. Single parents received less than married couples – a feature which is distinctive from the WFTC.

Single parents' employment trends were particularly interesting in the 1990s. This decade experienced the highest shift in the returns to education and skill. Employment rates of single mothers increased significantly from 60 to 73 percent, while employment was relatively stable for married women (Blundell and Hoynes, 2001). A large literature emerged attempting to explain the sharp changes in the female labor supply during the 1990s. The explanations provided by the literature included the expansion of the EITC in 1986, increases in the minimum wage,

Figure 1: EITC and WFTC schedules as of 2000. Source: Brewer et al. (2005)



and welfare reforms (Bingley and Walker, 1997; Card and Robins, 1996; Card et al., 1999; Dickens and Ellwood, 2001). In 1986, the largest EITC expansion began, where the maximum credit for families with one child (two children) reached $2206 ($3644), which implies the maximum amount was approximately doubled since the introduction of the program Hoffman and Seidman (2003). By 1996, the EITC became a central part of the US government's antipoverty programs, whose budget was 1.7 times higher than any federal spending on AFDC (Eissa and Liebman, 1996).

The eligibility requirements in 1996 were: first, the taxpayer needed positive income, which was the sum of wage and salary income, and business and farm self-employment income; second, the sum of all earned income should have been below a specified amount, which depended on the number of dependent children; third, all taxpayers needed a qualifying child: a child, grandchild, stepchild, or foster child under the age of 19 who lived with the taxpayer at least half a year prior to filing the taxes (Internal Revenue Service, 1996).

Eissa and Liebman (1996) is one of the earliest studies that investigates the labor force participation changes of single mothers following the expansion of the EITC in 1986 during the Tax Reform Act (TRA86). They argue that the expansion of the EITC led to dramatic shifts in the single mothers' employment rates relative to single women without children. Using the TRA86 as a "unique" event that changed the employment trends of single mothers, Eissa and Liebman (1996) follow a difference-in-difference (DID) identification framework to examine the labor force participation changes. This framework is useful in finding a "before and after" effect of the EITC expansion. They treat the change in TRA86 as a natural experiment which shifted the single mother's budget constraint. The three controls used were: (i) single women without children, (ii) single women with low levels of education, and (iii) single women who they predict would become eligible for the EITC based on exogenous characteristics (similar to

the propensity score approach). Eissa and Liebman (1996) use repeated cross-section data from the March Current Population Survey (CPS) in the US. They find that the labor force participation of single mothers increased by 1.9 – 2.8 percentage points relative to single women without children. The first control group was criticized because single mothers and single women without children exhibit very different employment patterns in the pre-treatment periods (Blundell and Hoynes, 2001; Blundell and MaCurdy, 1999). First, the DID conditions on the time effects and unobservables is unlikely to hold using repeated cross-sections. Second, single women without children were already working at full capacity during the 1980s (labor force participation was 95 percent) and therefore showed slower employment growth than single mothers prior to the EITC expansion in 1986. Eissa and Liebman (1996) do not show any evidence of the parallel trends assumption; therefore, it is difficult to assess whether this control is valid. Nevertheless, most of the literature in the early 2000s followed this approach.

Meyer and Rosenbaum (2001) examine the incentives of multiple federal program, including AFDC, Medicaid, Foodstamps, and EITC expansion in 1993, which was more generous to families with two or more children. They estimate a simple structural model of employment and find that the EITC accounted for over 60 percent of the changes in employment for single mothers during the 1990s. Their primary control group was single women without children. Similarly, Keane and Moffitt (1998) and Eissa and Hoynes (2004) find estimates of the same magnitude. These results primarily highlight that the EITC had stronger effects for single mothers with two or more children relative to those with one child due to the nature of the EITC expansion. Ellwood (1999) examines the EITC effects on marriage and cohabitation and finds no changes in marriage or cohabitation among low-skilled mothers, but some adverse effects on high-skilled mothers.

At least three concerns arise from the previous work in the EITC literature. First, the choice of the control (single women without children) might be problematic. Blundell and Hoynes (2001) and Blundell and MaCurdy (1999) argue that the pre-treatment employment patterns of this control group were not similar to that of single mothers because their employment rates were already near fully capacity. Hotz et al. (2006) note that other policies took place simultaneously with EITC. For instance, the Personal Responsibility and Work Opportunity Reconciliation Act (PRWOA) replaced AFDC in 1996, which affected single mothers and single women without children differently. Therefore, we cannot rule out the non-existence of other macroeconomic shocks that could cause changes in the differences between the two groups of women. Second, the March CPS underwent few compositional changes during the 1990s: the represented number of single mothers increased by 30.5 percent between 1984 and 1996 (Hotz et al., 2006). Meyer and Rosenbaum (2001) did not account for the "artificial" increase in the number of single mothers and thus their estimator is potentially upwards biased. Third, EITC claims have not been observed in the literature because of insufficient evidence from the datasets. Hotz et al. (2006) address the concerns above by restricting their analysis to California only because it experienced relatively less changes in other welfare programs than

other US states. Using a unique panel dataset under the DID identification strategy, they find that the EITC had an overall positive effect on the families that filed for the EITC in their tax returns. Their findings highlight the heterogeneity regarding the number of dependent children: the greatest effects were found among families with two or more dependent children relative to families with one child.

Overall, there is a general consensus in the literature that the EITC shifted the employment rates of single mothers positively. Eissa and Hoynes (2004) also examine the EITC effects on married couples. They find that in-work benefits had adverse effects on married women. Married women tend to substitute employment for leisure once their husband is eligible for the EITC allowance, while the effects on married men are moderately positive. This finding, initially suggested by Eissa and Liebman (1996) proves that in-work benefits may have created adverse effects when they are calculated upon family income and not individual income.

2.2 A Historical Review of In-Work Benefits in the UK

In-work benefits have a long history in the UK, predating the US EITC reforms. This anti-poverty strategy is believed to create unambiguous positive work incentives for families in the lower tail of the income distribution. Consequently, these benefits are targeted towards single parents, where single mothers comprise most single-parent households.

The first in-work benefits program in the UK was introduced in 1971, named Family Income Supplement (FIS). The FIS was reformed and replaced by the Family Credit (FC) in 1988, which was then reformed into the Working Family Tax Credit (WFTC) in 1999. I shortly describe the historical progress of in-work benefits in the UK and then compare it to the US EITC reforms.

Family Income Supplement

The FIS was an in-work benefit payable to low-income families with children, given that the household head was in part-time work. Part-time work was defined as 30 hours/week, but this limit was reduced in April 1979 to 24 hours/week for single parents (Sanctuary and Nurse, 1981). The allowance given to two-parent and single-parent households was increasing in the number of dependent children. Both the UK and the US transferred cash to families in need, but unlike the US, there was a universal transfer program named Child Benefit, which was worth £15 per week for the first child and £10 for subsequent children (Brewer, 2001). Child Benefit income was excluded from the calculation of the FIS payable amount, where the payable amount was half the difference between the family's income and the relevant limit[5]. The FIS was part of a package of benefits, including free school meals, free prescriptions, and dental treatment (see Dilnot et al. (1984) for more details).

[5]The limit in 1983 was £85.50 per week for a one-child family and an additional £9.50 per week for any subsequent child.

Although the FIS provided clear financial incentive to work, the combined effect of the entire package of benefits created ambiguous effects on a parents' willingness to work. Under the FIS, an eligible parent faced an overall implicit tax rate of 105 percent (Blundell and Hoynes, 2001). This implied that the individual would be better off not working than working. After the FC reforms, the implicit tax rate reduced to 97 percent, creating better financial incentives to work.

Family Credit

In 1988, the FIS fully reformed into the FC, which was designed to increase the generosity of the FIS and remove the tax rates in excess of 100 percent. The structure changed in 1992, where the eligibility requirement of "part-time work" was no longer 24 hours/week, and was set at 16 hours/week. The FIS had a minimum requirement of 30 hours/week for two-parent households and 24 hours/week for single-parent households, while the FC enforced a 16 hours/week threshold for all parents and added a supplementary credit for those working 30 hours/week to avoid any adverse effects on willingness to work.

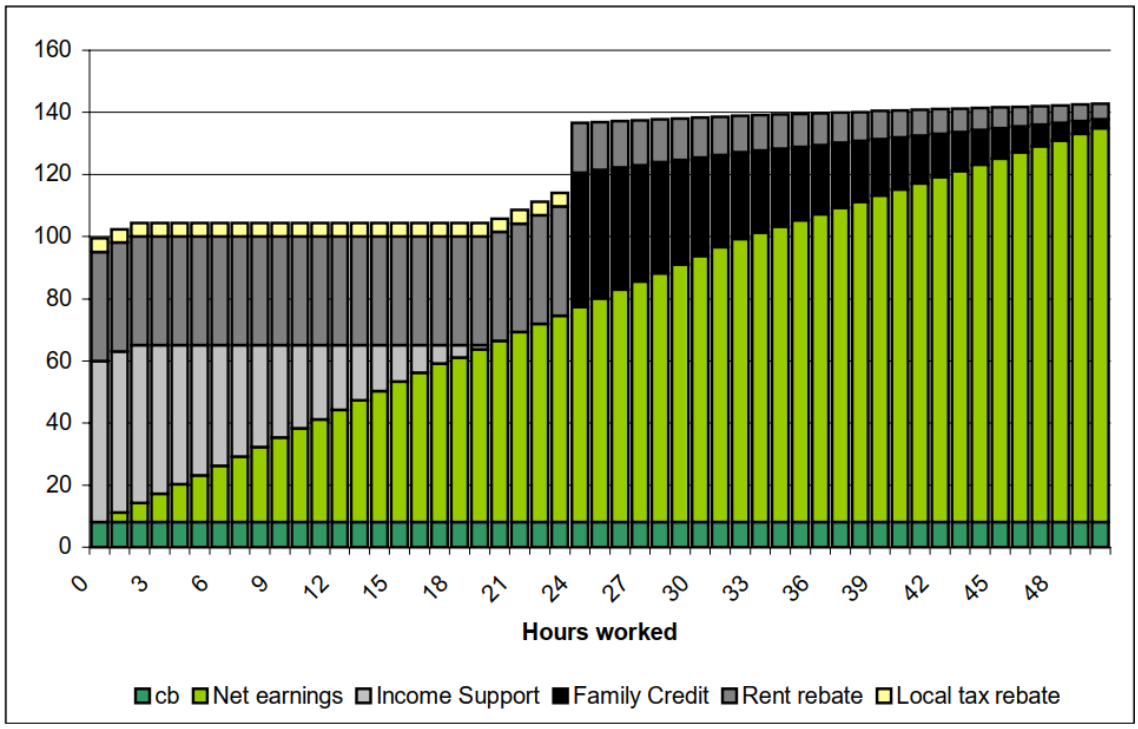
The payable amount under the FC depended on the paid hours of work, the number of eligible children, and household income. Income above a certain threshold (£79.00 per week) reduced the credit at a rate of 70 percent (the phase-out region on the credit schedule). Unlike the US, the FC income was counted when calculating other benefits, including Housing Benefits (HB). The HB were withdrawn at a rate of 65 percent when income reached a certain threshold. This high withdrawal rate interfered with the financial incentives of the FC. When household income fell below a threshold, all rental payments were paid by the UK government. However, when these households were eligible for the FC, income increased, and HB decreased with a high slope, which lowered the overall income of the household. Hence, the FC may have create adverse effects on the individual's willingness to work.

The interaction between FC and HB is shown in Figure 2. In 1991 (figure 2a), an eligible single parent would start receiving FC income when working at least 24 hours/week. Child Benefits (CB) were available to everyone, while Income Support (IS) and local tax rebates were available to individuals not working less than 24 hours/week. Rent rebate (HB) was significantly reduced by 65 percent once the individual was eligible for the FC because the FC was treated as income when calculating HB. Given that eligibility for the FC diminished the extra income for HB, this may have reduced the financial incentive of the FC to "make work pay" (Giles et al., 1997). In 1992 (figure 2b), the FC lowered the eligibility to 16 hours/week, but the withdrawal rate of the HB remained the same. Therefore, the increase in income when an individual became eligible for the FC was minimal, which may have created adverse effects on willingness to work.

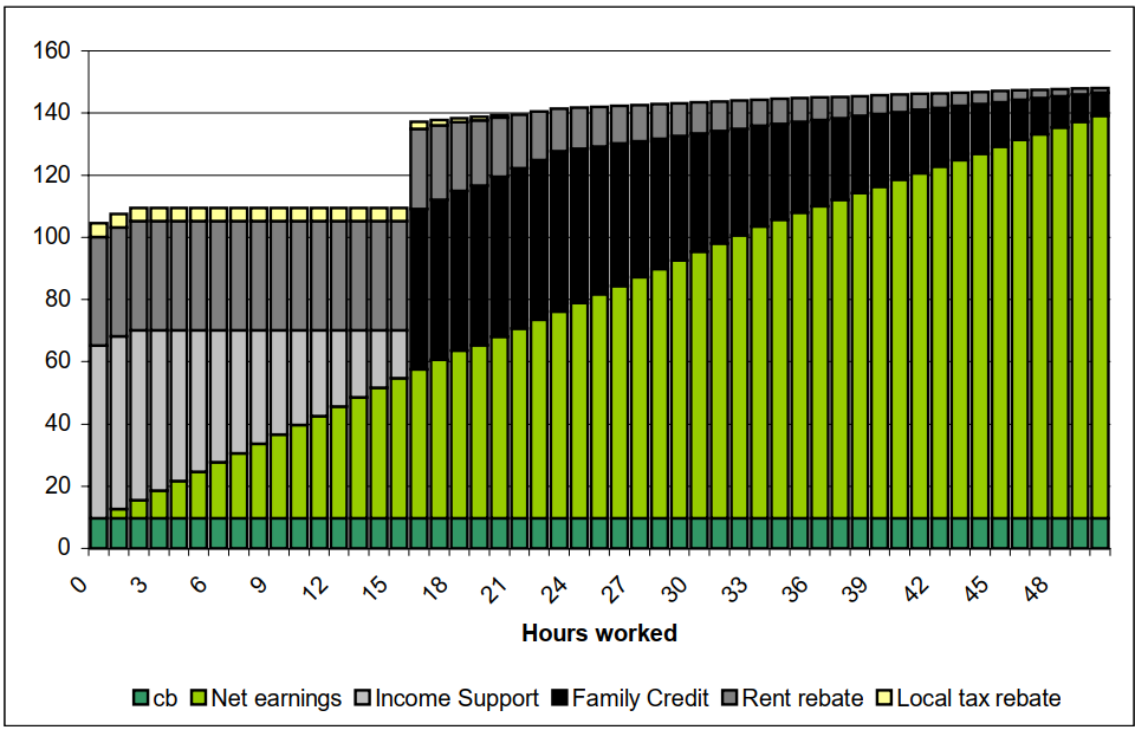
Working Family Tax Credit

The WFTC replaced FC in October 1999, but was fully phased in from April 2000. The WFTC was introduced along with other policies, including the New Deal for Lone Parents

Figure 2: Net income of a single mother before and after FC. Source: Blundell and Hoynes (2001)



(a) Single parent's budget constraint in 1991, working 3.00 per hour.



(b) Single parent's budget constraint in 1992, working 3.00 per hour.

(NDLP) and the National Minimum Wage (NMW), all of which played a crucial role in the government's antipoverty programs. The WFTC was the main in-work benefit program in the UK until April 2003, which was then reformed into the Working Tax Credit (WTC). By the end of its duration, the WFTC covered nearly 1.4 million families and 2.7 million children, which accounted for a 70 percent increase in its coverage compared to FC (Francesconi and Van der Klaauw, 2007).

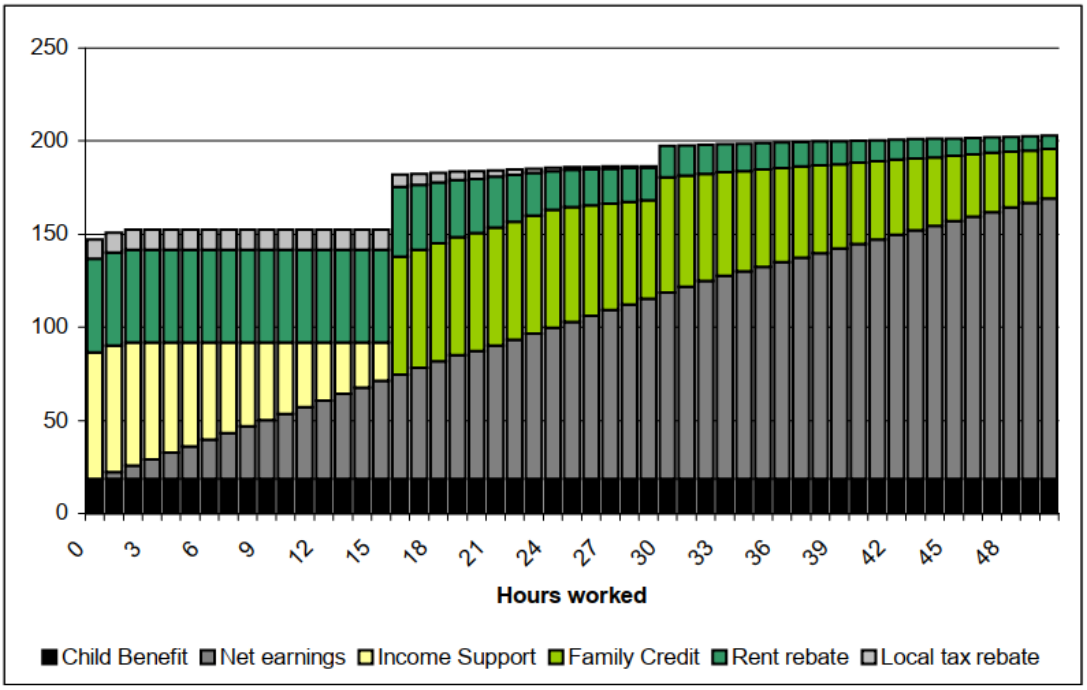
A family was considered eligible if they met three basic requirements. First, at least one parent in a household should have worked 16 hours or more per week. Second, the parent(s) needed at least one dependent child, which was considered a child, grandchild, stepchild, or foster child until the age of 16 or 19 if they were enrolled in full-time education. Third, the household's savings should have been less than £8,000 to be considered eligible, while savings between £3,000 and £8,000 reduced the credit. The amount of credit depended on weekly earnings, hours worked, the number of qualifying children, and household savings. The WFTC was more generous than the FC in four significant ways. First, the WFTC credits for families with more children and younger children[6] were 34 percent higher than the FC. Second, the income threshold from which the allowance was withdrawn increased by 46 percent, from £62.25 per week under FC to £91.45 per week under WFTC. Third, the withdrawal rate when the parents' income was higher than the threshold was smaller (55 percent withdrawal rate compared to 70 percent under the FC). Fourth, the WFTC incorporated new childcare credit that covered 70 percent of actual childcare costs up to a higher amount relative to the FC (Blundell and Hoynes, 2001). Twice as many families became eligible under WFTC compared to FC. The government spent nearly £5 billion per year on the WFTC, which was a 60 percent increase from the FC's budget. Figure 3 shows the budget constraint of a single parent before and after the introduction of the WFTC. It is evident that the WFTC payments were much higher, essentially eliminating the rent rebate (HB) in exchange for the WFTC when the parent worked 30 hours or more per week.

In 1994, the FC allowed childcare costs to a maximum of £60 per week. However, these costs were not calculated when calculating the family income for other supplementary services (like the HB). The WFTC covered 70 percent of childcare costs up to £100 per week for one child and £150 per week for two or more children, implying that eligible families would receive £70 and £105 a week, respectively (Francesconi and Van der Klaauw, 2007). Additionally, the FC was administered by the Benefits Agency, while the WFTC was administered through the employer via the Inland Revenue to reduce the stigma of claiming benefits through bureaucratic procedures.

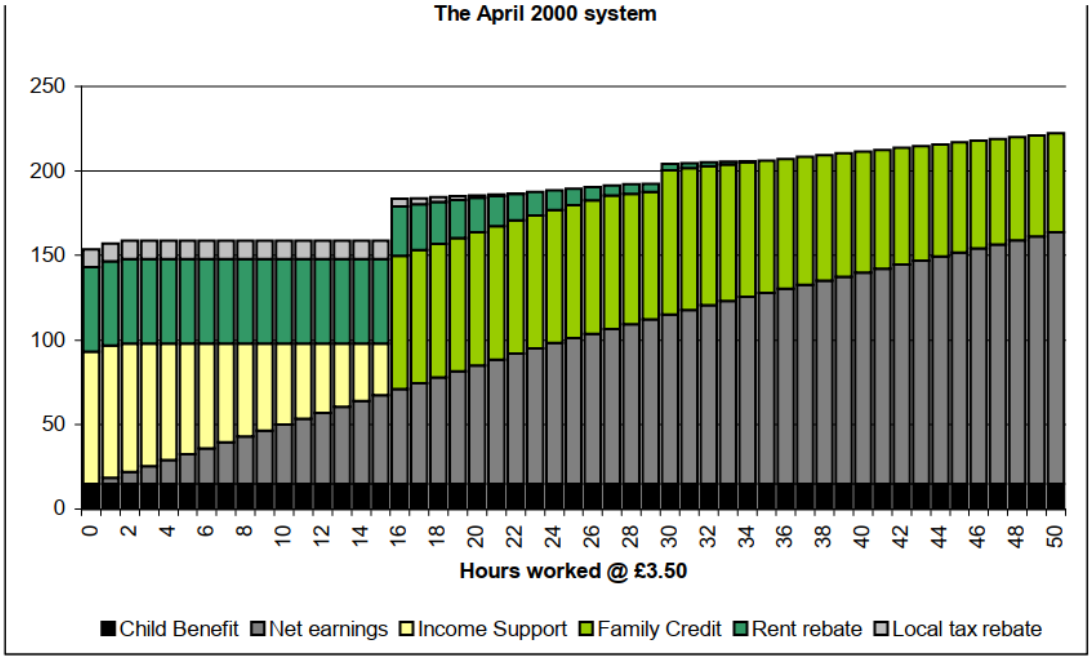
To assess the financial incentive of the WFTC, figure 4a shows the WFTC and FC payment schedules, and figure 4b the budget constraints of a single parent under both WFTC and FC. In two-parent households, the WFTC required both partners to work more than 16 hours/week to receive the childcare credit. Consequently, this avoided the adverse effects of previous policies

[6]Children aged 0 – 10 years old were considered young.

Figure 3: Single parent's budget constraint before and after the WFTC. Source: Blundell and Hoynes (2001)



(a) Single parent's budget constraint in 1998, working £3.50 per hour, 2000 prices.



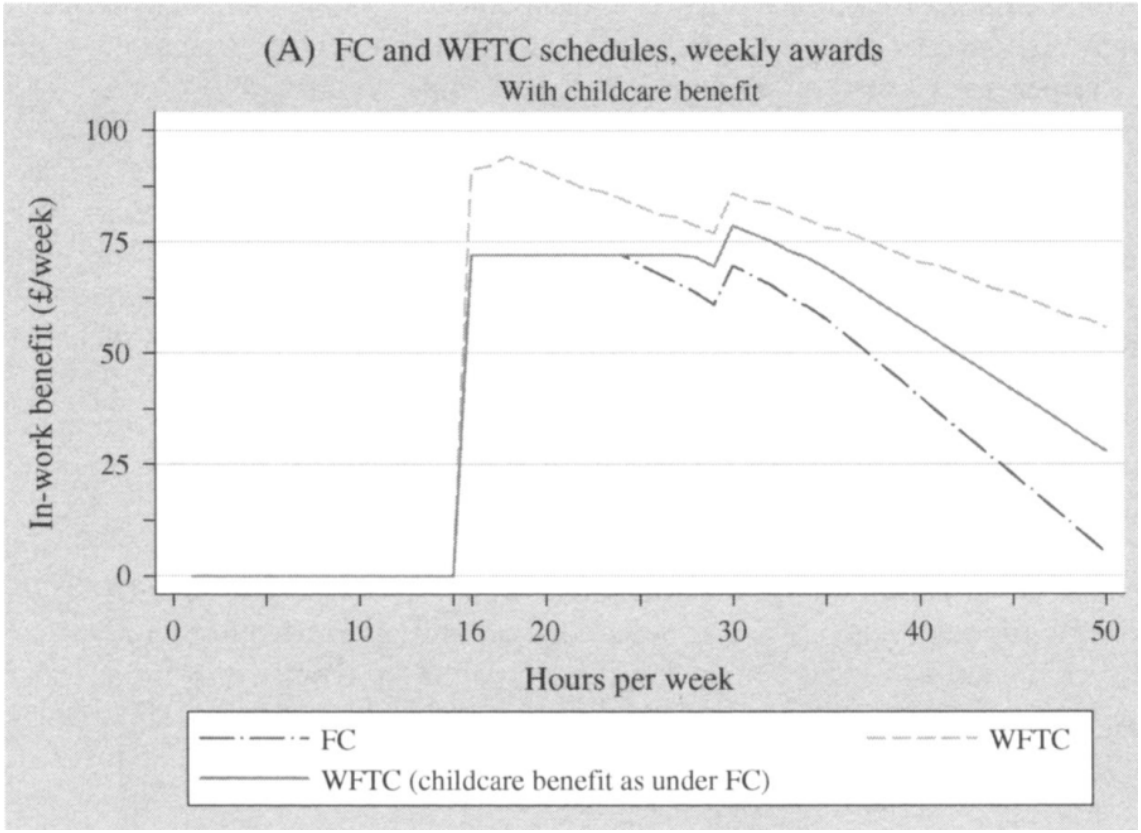
(b) Single parent's budget constraint in 2000, working £3.50 per hour, 2000 prices.

(FIS and FC). Unlike the EITC schedule, there are two regions in the WFTC schedule. When an eligible parent reached the maximum amount of £91.45 a week, the credit was withdrawn at a rate of 55 percent, and the parent became ineligible when weekly earnings were at least £385 per week.

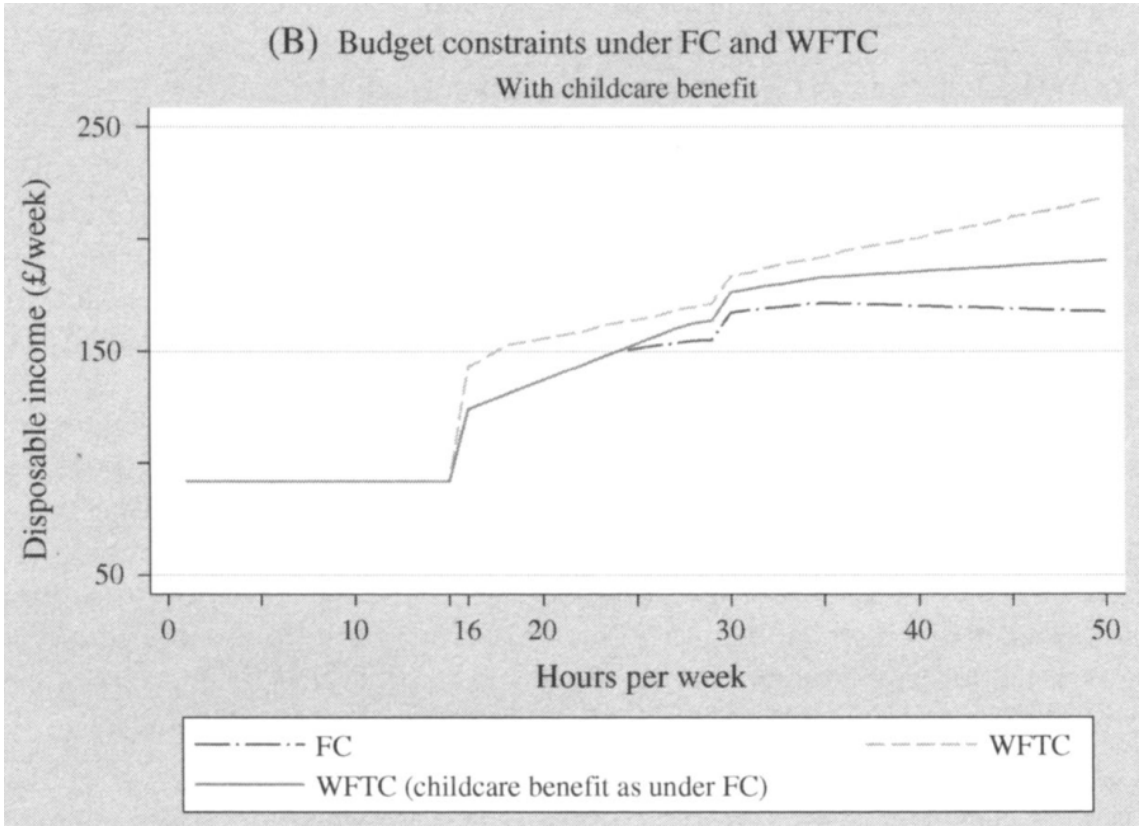
The WFTC schedule was unambiguously higher than the FC and the WFTC under FC child benefits. Under the newly reformed WFTC, a single mothers' budget constraint strongly dominated the FC budget constraint once the mother became eligible. Figure 4 suggests that WFTC shifted more single mothers into part-time employment (16 hours/week), but the effect on paid hours of work depends on the income and substitution effect and is difficult to predict (Brewer, 2001). Figure 4 also suggest that the childcare costs covered by the WFTC are advantageous to the FC childcare coverage. Hence, the WFTC reduced the burden of childcare costs, and potentially induced single mothers to use childcare services with the intent to work more.

To obtain a better understanding of the WFTC reform, let us consider a single mother with one child aged 5 years old who is working more than 16 hours/week, pays £60 per week on childcare, and earns £150 per week. Under the FC regime, this mother would receive £56.80 credit per week, while under the WFTC, the same mother would receive £81.15 per week (in 1999 prices). This difference reflects an approximate 43 percent increase in income, which is quite substantial for a low-income single mother. For the entire group of single mothers, the overall increase in income would be 20 percent, while at least a quarter of the single mothers would see a 50 percent increase in their weekly income (Francesconi and Van der Klaauw, 2007). Nevertheless, the income supplemented via the WFTC was decreasing in HB.

Figure 4: FC and WFTC schedule and budget constraint. Source: Francesconi and Van der Klaauw (2007)



(a) Single parent's budget constraint in 1991, working 3.00 per hour.



(b) Single parent's budget constraint in 1992, working 3.00 per hour.

2.3 The Link Between the UK and US In-Work Benefits

Although the in-work benefits system in the UK predates the one in the US, the WFTC was introduced because the UK government wanted to incorporate the advantages of the EITC to alleviate poverty from the low-paid families (Hansard, 1997). Both the EITC in the US and the WFTC in the UK supported working families and aimed to reduce child poverty, and there was a strong political connection between the two programs. Nevertheless, as of June 2000, there were substantial differences between the two in four main aspects (Brewer and Gregg, 2003).

Eligibility

The eligibility for the WFTC depended on paid labor hours (where the parent worked 16 hours or more per week), the presence and the number of dependent children (either under 16 or under 19 years old if they were enrolled in full-time education), and household savings (savings over £3,000 reduce the credit while savings over £8,000 removed eligibility). Under the WFTC, couples were assessed jointly; however, there was no need for parents to be married. There was a possibility for adults without dependent children to file for the WFTC if they proposed alongside an integrated child credit. The eligibility for the EITC depended on three requirements. First, the adults must have positive earnings in the past year and an annual investment income under $2,350. Second, similar to the WFTC, married parents were assessed jointly, while single parents could file separately. Third, parents needed to have an eligible dependent child, which was either under 19 or under 24 and enrolled in full-time education, or permanently and totally disabled. It is important to note that when a child qualifids as a dependent child for two single parents, only the parent with the highest income could apply for the EITC.

Structure

Figure 1 presents the EITC and WFTC schedules as of June 2000. The WFTC awards were assessed on weekly earnings, and it was calculated based on the parents' past four paychecks. Then, the WFTC was paid biannually, regardless of any income changes. The childcare tax credit was supplementary to the WFTC payments. There was a basic credit for working between 16 to 30 hours/week; and an additional £11.25 if the parent worked more than 30 hours/week. Additionally, payments increased when the parent had two or more dependent children – the extra allowance was approximately £26.00 for children aged 16-18 years old, and up to £40.00 for younger children. On the other hand, the EITC payments were paid annually and operated as an annual tax rebate. The maximum yearly amount depended on the number of eligible children: $2,353 for families with one child, and $3,888 for families with two or more children.

The WFTC had no phase-in region, while the phase-out (withdrawal rate) was 55 percent once the weekly earnings reached £91.45 per week. On the other hand, the EITC had a phase-in region which also depended on the number of dependent children. The credit was 34 percent

of income for families with one child, and 40 percent for families with two children. There was a maximum credit region, which was approximately $10,000 for families with two children and $8,000 for families with one child. The phase-out rate also depended on the number of children: 16 percent and 21 percent for families with one child and two children, respectively.

Interaction with other benefits and taxes

Both EITC and WFTC payments were calculated based on the parents' net income; however, the definition of "net income" differed between the two benefits. For the WFTC, net income was defined as income after tax and national insurance, while the EITC used gross earnings before tax. Multiple benefits were excluded from the definition of income when calculating WFTC and EITC payments. For the WFTC, child benefits, statutory maternity pay, attendance allowance, maintenance payments, HB, and council tax benefits were not counted as income and were paid separately to the WFTC. However, for the calculation of HB and council tax benefits, WFTC payments were counted as income and could therefore dampen the other payments. For the EITC, Temporary Assistance for Needy Families and Food Stamps were not counted in the gross income. Federal law prohibited EITC to be counted as income when calculating Medicaid, Supplemental Security Income, Food Stamps, and AFDC.

Assessment and payment mechanism

The WFTC payments were assessed on the frequency of the parents' income. For weekly paychecks, the assessment period was based on the past four paychecks; for bimonthly payments, the assessment was based on the past 4 paychecks; while for monthly paychecks, the assessment was every four months. The payments were given every 26 weeks and it was given through the wage packet from the employer. The employer was then reimbursed by the Inland Revenue. When couples jointly filed for the WFTC, they decided amongst themselves who received the payment; otherwise, the Inland Revenue gave it to the main earner. Unlike in the UK, the US government supplemented EITC through the tax system. Therefore, the EITC payments were annually based on gross income from the year prior to when the families filed their taxes (which is typically by April of each year). However, a maximum of $1,418 could be paid in advance for parents whose federal income tax was withheld from their wages. Similar to the WFTC, couples decide amongst themselves who received the EITC payment when they file for taxes.

2.4 The WFTC Literature

There is a large literature that analyzes the effects of in-work benefits on the labor market effects of single mothers. The WFTC was one of the first reforms of the Labor Government, whose stated intention was to eradicate child poverty through cash transfer programs.

By the beginning of the new millennium, evidence regarding the introduction of WFTC was rapidly increasing. Blundell and Hoynes (2001), Gregg and Harkness (2003), and Paull et al. (2000) attempted to predict the employment changes after the WFTC reform. They predicted that employment should rise by approximately 2 percentage points for single mothers, while employment should fall for married mothers. Brewer (2001) published the first paper that studied the effect of the WFTC on employment. He focused on the labor market effects on single parents, using single adults without children as the control group. His results suggest that the WFTC increased the employment rate of single parents by 1.4 percentage points; however, his dataset contained information only until Summer 2000, while the WFTC was fully implemented in April 2000. Therefore, their estimates might suffer from large standard errors and may be statistically imprecise. Gregg and Harkness (2003) used data from the Labor Force Survey from 1998 to 2002 to assess the impact of the package of policies implemented alongside the WFTC. They used a DID identification strategy to recover the average treatment effects (ATE). Although the DID framework does not find individual policy effects, the conditional DID gives powerful evidence on the behavioral effects on the labor supply of single mothers (Ellwood, 1999). Using single women without children as controls, they find that the WFTC positively shifted the employment rates of single mothers by 5 percentage points. Additionally, they check for heterogeneous treatment effects regarding the age of the youngest child and education of the mother, and found that the WFTC was more effective for single mothers with younger children; however, not for the least educated mothers. Given their rich dataset, they were able to divide the single mothers post-policy into incumbents (those who were employed pre-policy) and entrants (those employed only post-policy) and find that the entrants worked fewer hours than incumbents because they had a higher "distaste" for work and were only induced into employment because of the WFTC.

Blundell et al. (2005) is the most comprehensive study of WFTC on labor market outcomes. They estimate a probit model of employment, using a DID identification strategy and repeated cross-section data from the Labor Force Survey from 1996 to 2003. They provide an extensive discussion on the choice of the control, where they agree with Gregg and Harkness (2003) on using single women without children. However, the parallel trends assumption was not satisfied because the employment of single mothers was seemingly increasing at a faster rate than single women without children starting from 1998 (one year prior to the introduction of the WFTC). Blundell et al. (2005) noted that the Income Support was reformed prior to the WFTC. However, Income Support was given to unemployed individuals; hence, it incentivized people to not work. Therefore, the effect of the WFTC would be understated. Their estimate suggests a small and positive effect for single mothers, and the effects were diminishing over time. Similar to Gregg and Harkness (2003), they find heterogeneous treatment responses regarding the number of children and the children's age.

Leigh (2005) makes use of a fifteen-month panel dataset to track the same parents pre- and post-WFTC to compare their labor market responses. Their dataset contains months from 1999

to 2000, implying that they measure the immediate effect of the WFTC. Using the entire eligible group of recipients (adults with children) as their treatment group and non-eligible adults as their control, they find that the overall effect of the WFTC was a 1 percentage point increase in employment and 1.1 hour increase in paid work. This policy also increased earnings by 4 percent and potentially reduced the number of people who faced serious health problems. Contrary to the predictions by Blundell and Hoynes (2001), Gregg and Harkness (2003), and Paull et al. (2000), Leigh (2005) did not find any heterogeneity across household structure, implying that married mothers did not behave differently from single mothers. However, their estimates do not provide information about the long-term effects of the WFTC nor the persistence of the effects due to their short panel.

Francesconi and Van der Klaauw (2007) document the heterogeneity in the employment across single mothers. They use a novel panel dataset from the British Household Panel Survey from 1998 to 2001 to estimate the persistence and entry probabilities of employment of single mothers. Differently from previous papers, they use a difference-in-difference-in-difference (DDD) identification framework because their parallel trends assumption was not satisfied and therefore accounted for pre-reform trends of single mothers that differed from single women without children. Their findings offer insightful evidence of the effects of the WFTC. First, they find that most of the increase in employment was driven by the entrants (single mothers working only post-WFTC), while for incumbents (those working pre-WFTC) they found no effect on paid hours of work. They conclude that the main driver of the increase in employment was the substitution effect, meaning that the WFTC induced non-working single mothers to work part-time (at least 16 hours/week) because of the benefits. Second, they find significant heterogeneity for single mothers whose children were under 5 years old, but no heterogeneity from the number of children. Third, their findings suggest that single mothers were significantly less likely to cohabitate or marry, and there is weak evidence that they were less likely to have children again. Fourth, they find that the treatment effects could be driven by the generous Childcare Tax Credit component of the WFTC, which led to an increase in the use of paid childcare and alleviated the childcare burden of single mothers. Brewer and Browne (2006) also find evidence that financial incentives affect fertility. They argue that the WFTC had potentially ambiguous effects depending on the marital status of the mother. They find that the WFTC did not increase fertility for single mothers but did significantly increase fertility for married mothers. Single mothers faced a negative opportunity cost of having a child and therefore demanded children less when they had the financial incentive to work. Married mothers, on the other hand, faced a positive opportunity cost and demanded more children with the WFTC. Additionally, they found evidence that there was a higher response for mothers who were having their first child than those with subsequent births.

2.5 Identification Issues from the WFTC Literature

In this section I compare the five main studies in the WFTC literature: Brewer (2001), Leigh (2005), Gregg and Harkness (2003), Blundell et al. (2005), Francesconi and van der Klaauw (2004). Overall, the estimates from each paper seem to vary substantially and this depends on the assumptions they use for their identification approach and their outcome variables. This literature seemingly agrees that the WFTC increased the proportion of single mothers in the labor force and increased the hours of paid work. Nevertheless, there are substantial differences in the approaches each study takes.

Time period

Leigh (2005) and Brewer (2001) use a panel dataset from 1999 and 2000, Francesconi and van der Klaauw (2004) use data from 1997 to 2001, Gregg and Harkness (2003) study the period 1998 to 2002, while Blundell et al. (2005) use data from 1996 to 2002. Although the differences in the time periods seem small, they are significant for two reasons: (i) the richness of the pre-WFTC data allows the researcher to check for the parallel trends assumption and gives more correct estimates, and (ii) the WFTC further expanded during 2000 and 2002, which means that the largest effect would be during 2002 Brewer and Browne (2006). Therefore, Leigh (2005) and Brewer (2001) measure the immediate effect of the WFTC, which may be an understatement of the true, long-term effect.

Pre-WFTC Parallel Trends Between Single Mothers and Single Women Without Children

Brewer (2001), Leigh (2005), Gregg and Harkness (2003), Blundell et al. (2005), Francesconi and van der Klaauw (2004) use the DID identification strategy to estimate the causal effect of the WFTC on the labor market responses of single mothers. All four papers define their treatment group based on eligibility (having children); therefore, their main treatment group were single mothers, and their main control were single women without children. The DID approach does not identify individual changes due to the treatment but estimates the average treatment effect of the individuals. The control is chosen such that they experience the same labor market shocks that the treatment group also does, except for the WFTC. Prior to the WFTC, other welfare programs were changed or introduced that affected the labor market behavior of the population. An ideal control would be a group of individuals that were affected by these other programs and not the WFTC. Unfortunately, the other welfare policies may have affected single mothers more than single women without children because single mothers tend to be younger, less educated, and poorer. All studies found that the parallel trends assumption was *not* satisfied.

In April 1999, a National Minimum Wage (NMW) policy was introduced for the first time. This policy affected low-skilled individuals more than high-skilled, implying that single mothers were more affected by the NMW. The NMW increased the gap between in-work and out-

work outcomes Brewer and Browne (2006); hence, the NMW potentially affected single mothers more than single women without children. Additionally, the effects potentially manifested months later and interacted with the WFTC effects. This implies that the estimates found by the literature could be overstating the true WFTC effect. However, Dickens and Manning (2004) show that this was not a problem because the effects of the NMW occurred within the fist two months of its introduction, and the effects were rather small.

The New Deal for Lone Parents (NDLP) was introduced in 1998. This welfare program encouraged its participants to attend Work-Focused Interviews under the Jobcentre Plus program. Its aim was to train single parents to become better interviewees, which could increase their chances of finding (better) jobs. However, these programs were voluntary and did not ensure the participants that they would find jobs. Gregg and Harkness (2003) and Brewer and Browne (2006) argue that only a small fraction of single mothers was participating in NDLP; therefore, the bias resulting from this program would likely be small.

A variety of explanations were given to justify the violation of the parallel trends assumption. Gregg and Harkness (2003) argue that the change in the employment of single mothers would have occurred irrespective of the WFTC. Blundell et al. (2005) and Leigh (2005) argue that the increase was temporary and would not have continued after 1998, while Francesconi and Van der Klaauw (2007) attribute this to the "anticipation effect". I provide a more extensive discussion on this issue in section 6.

3 A Theory of Labor Supply

The WFTC increased the single mothers' income the more hours she worked. The budget constraint in figure 4b shows that the WFTC shifted the single mothers' opportunity set compared to the FC. Given that the WFTC was paid based on labor hours and through the employer, I classify it as labor income. In this section, I present a simple framework of labor supply under the circumstances of a wage change, as discussed by Borjas (2020).

Consider a wage increase holding all non-labor income constant. A working individual faces the ultimate trade-off between consumption C and leisure L when deciding the amount of hours/week they should work, both of which are normal goods[7]. Irregardless of the workers preferences, the budget constraints shifts outwards when there is a wage increase, implying that the worker has more income and can afford to buy more goods. However, the effect of a wage increase on preferences for C and L, and ultimately hours of work, is ambiguous. Figure 5 shows the decomposition of the income and substitution effects when one or the other dominates. It is worthy to note that both effects play a role in determining the change in hours of work and in some cases could cancel each other out.

Figure 5a shows the case when the income effect dominates the substitution effect. Initially, the worker is at equilibrium P, working 40 hours/week and spending 70 hours/week on leisure activities. An increase in the real wage rotates the budget constraint around the endowment point **E** from *FE* to *GE*. After the wage increase, the worker is at equilibrium point R, working 35 hours/week and spending 75 hours/week on leisure. The movement from P to R can be decomposed into two effects:

> **Stage 1:** P $\rightarrow$ Q, the worker "feels" that their budget constraint parallel-shifted to DD, where the tangency point with the indifference curve U_1 is Q. The worker's initial response to a wage increase is to work only 25 hours/week and spend 85 hours/week on leisure. This is the *income effect*.
>
> **Stage 2:** Q $\rightarrow$ R, because the returns to working are higher, the opportunity cost of not working is also higher, implying that the price of leisure is more expensive. This encourages the worker to reduce their demand for leisure and increase their demand for work and other consumption goods. This is the *substitution effect*, which causes the worker to substitute leisure with labor.

In this setting, the income effect dominates the substitution effect, leading to a reduction in labor hours. Figure 5b shows the case when the substitution effect dominates the income effect. The movement form P to R is also decomposed into two effects:

> **Stage 1:** P $\rightarrow$ Q, is the same income effect as in figure 5a, which arises from the additional income the worker receives by the wage increase.

[7] By normal goods I imply that the demand for C and L is increasing with income.

Stage 2: Q $\rightarrow$ R, which is the substitution effect. This symbolizes a move away from leisure time towards labor hours and consumption of other goods. This worker devotes more time to work because they have a higher taste for labor and a higher distaste for leisure. Leisure is more expensive for this type of worker.

In this setting, the substitution effect dominates the income effect, leading to an increase in labor hours. Whether the income or substitution effect dominates depends on the type of worker. To analyze the income and substitution effects after the introduction of the WFTC, I would need data on the mothers' preferences for consumption and leisure. In absence of this data, I can only argue under which conditions one of the two cases may occur.

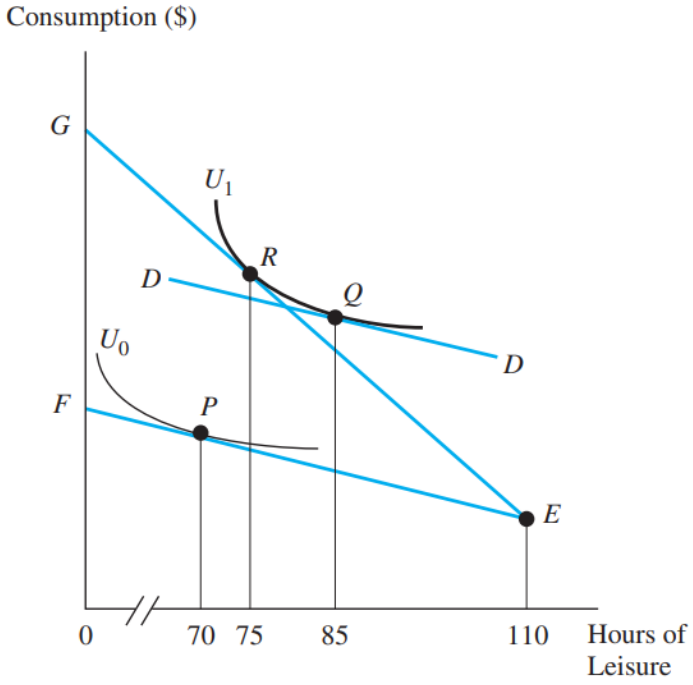
Case 1: Substitution effect $>$ Income effect. In this case, the single mother has a higher preference for work, or leisure is too expensive for her. This might potentially be the case for single mothers with lower labor hours who had lower income, less children, or older children. For these mothers, there was more incentive to work and having less/older children gave them more "freedom" to work under the WFTC incentive.

Case 2: Income effect $>$ Substitution effect. In this case, the single mother prefers leisure over work and reduces her labor hours. This might be the case for single mothers who were working more hours, had more income, and those who had younger children and/or more children.

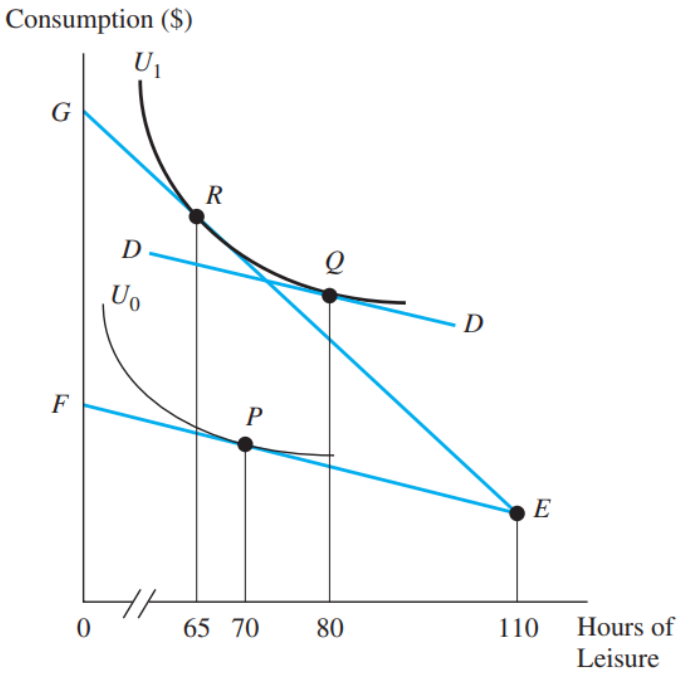
Eissa and Liebman (1996) and Meyer and Rosenbaum (2001) in the case of the EITC and Gregg and Harkness (2003) and Francesconi and van der Klaauw (2004) in the case of the WFTC all find that in-work benefits increase employment more for mothers with younger children. However, as microeconomic theory suggests, the effect on employment might not be the same as the effect on labor hours.

The literature agrees that the WFTC made employment unambiguously more attractive. However, the effect on the hours of work is less clear because it depends where the person would choose to be on the pre- and post-WFTC budget constraint (Meyer and Rosenbaum, 2001). Along with the income and substitution effects, there is an additional effect that is specific to tax credits or other government programs: the negative substitution effect. The WFTC withdrawal rate began immediately when the mother worked 16 hours or more per week, with a small jump in payments when the mother worked 30 hours/week, and then the withdrawal continues until the mother was no longer eligible. This phase-out region of the WFTC schedule causes the negative substitution effect: the returns to working were lower; therefore, single mothers substituted labor for leisure hours, resulting in a reduction in hours of work. Meyer and Rosenbaum (2001) predicted that the EITC in the US would lead to a small reduction in labor hours because of the income effect and negative substitution effect which would dominate the positive substitution effect. Additionally, because the WFTC was counted as income when calculating housing benefits (HB), it may have led to a reduction of labor hours for those single mothers who worked 16 hours or more per week.

Figure 5: Decomposing the impact of a wage change into Income and Substitution Effects. Source: Borjas (2020)



(a) Income Effect dominates



(b) Substitution Effect dominates

4 Data

4.1 The Labor Force Survey

The Labor Force Survey (LFS) is a representative survey of UK households, with sample sizes of nearly 60,000 households in each quarter[8](Office for National Statistics, 2002). It is a unique source of information regarding information on a variety of topics, including employment and unemployment, together with a wide range of information related to occupation, training, hours of work, and household composition and characteristics for all household members older than 16 years old. It includes all adults (16 years old or older) that reside in private households, hospital accommodation, and student halls. The LFS was first conducted in 1973. Between 1973 and 1983, the survey was carried out once every two years, while between 1984 and 1991, the survey was annual. From 1992, the quarterly data (QLFS) was made available. In 1994, the QLFS included Northern Ireland; hence, providing information for the whole of the U.K[9]. In this thesis, I use the QLFS.

In each quarter, the QLFS comprises of a "core" of questions, together with some "non-core" questions which are spontaneous from quarter to quarter. The first part of the questionnaire contains questions on the respondent's household, family structure, basic housing information and demographic details. The second part of the core questionnaire consists of questions on economic activity, education, and health. Detailed questions on income have been included since 1993. The QLFS is advantageous for employment-related research in many regards. First, the survey asks very detailed questions on the households' labor supply decisions; second, the large sample and the repetitiveness of the survey allows more variation in the sample; and third, the structure of the LFS can be used for cross-sectional, repeated cross-sectional, and for longitudinal analysis (for linked 5 quarters).

Nevertheless, there are some important limitations to the QLFS. There are "missing" (No response) and "Does Not Apply" (DNA) values in all variables, and the QLFS online guideline does not precisely explain the reason why these values are missing or DNA for some observations (UK Data Archive, nd). These values create issues for two reasons. First, DNA answers take negative values and significantly distort the way the data appear in the analysis. Second, when labor supply data are missing for non-workers, it leads to the selection problem, which is important to account for if a researcher wants to estimate the elasticity in labor supply functions (Heckman, 1993). There is no consensus in the econometric literature about particular solutions to the selection problem in the labor market, and each solution proposes different assumptions about the unobserved terms for the counterfactual[10] (the wages and labor hours non-workers would work if they chose to work). As per the advice of Professor Menzel[11], I

[8]The first quarter begins January - March of a given year, and one year consists of four quarters.

[9]The QLFS is publicly available **here.**

[10]I distinguish *control* and *counterfactual* here. The control group is what I choose as the untreated individuals, while the counterfactual group is the treated group in the absence of the treatment.

[11]Andreas Menzel, Ph.D. Assistant Professor at the Center for Economic Research and Graduate Education -

apply the following strategy: for each variable, I replace all "DNA" with "missing" values. Then, for all variables that contain at least one missing value, I create a dummy variable which equals to 1 if a value is missing for that variable, and 0 otherwise. I then control for the missing values in my analysis by including the dummy variables as covariates.

I construct a repeated cross-section dataset with 22 quarters from the QLFS, beginning from Spring 1996 to Winter 2002. The WFTC underwent significant changes and introduced new components after 2002; hence, observations after 2002 are not included. My main outcome variable, paid labor hours, was not collected during 1999 by the QLFS for unknown reasons (Office for National Statistics, 2015). Additionally, Blundell et al. (2005) suggest that observations from the first quarter in 2000 should be dropped to allow both for the full phase-in of the WFTC and to leave enough time for eligible parents to adjust their labor market behavior. Therefore, I omit 4 quarters from 1999 and one quarter from 2000. My sample includes all employed women who are single, divorced, or widowed. It includes more than 411,000 women (19,000 in each quarter), 44 percent of which are single mothers (the treated group) and 56 percent of which are single women without children (the control group). I define a single mother as any woman in my sample who has more than one dependent child[12].

4.2 Outcome Variable

My outcome variable is *paid labor hours*. This variable takes the total paid hours actually worked by respondents in their main job during the reference week, including paid overtime too. It applies to all respondents who are employees, self-employed, and those on employer-based government schemes, and if respondents are away from their job, or off sick for the week in question, then this variable takes the value "0". Table 1 shows mean changes of paid hours by different quantiles.This table indicates that the differences of paid labor hours between single mothers and single women without children are different across quantiles. This simple difference-in-difference approach indicates the presence of heterogeneous treatment effects, with a positive effect in the bottom quantiles, and essentially no effect on the top quantiles. The summary statistics for paid labor hours are shown in table A1. Pre-WFTC, single women in my sample worked 24 hours/week. The maximum value is 97 hours/week, although very few observations take that value. The standard deviation is quite high, implying that many single women lie on different locations of the paid labor hours distribution. This is another indication of the existence of heterogeneity, implying that looking for mean impacts of the WFTC would not be insightful. For illustrative purposes, I also show the share of single women working more that 16 and 30 hours/week in my sample. Approximately 27 percent of women work more than 16 hours/week, while only 16 percent work 30 hours/week or more. However, this table is not informative of the labor supply differences between single mothers and single women

Economics Institute (CERGE-EI).

[12]The QLFS defines a dependent child according to the same definition as the Inland Revenue.

Table 1: Distribution of paid hours

	Pre-WFTC		*Post-WFTC*		*Difference SM - SW*		
	SM[a]	SW[b]	SM	SW	pre-WFTC	post-WFTC	DID
Quantile							
10th	4.54	11.74	5.87	12.04	-7.2	-6.17	1.03
20th	8.51	23.87	10.79	23.72	-15.35	-12.72	2.63
30th	14.15	30.76	15.26	30.77	-16.61	-15.51	1.1
40th	19.05	35.18	19.05	35.17	-16.13	-16.12	0.01
50th	23.27	37.69	23.22	37.66	-14.42	- 14.44	0.02
60th	29.5	39.68	29.48	38.66	-10.18	- 9.18	1
70th	35.56	40.15	35.6	39.74	-4.59	-4.14	0.45
80th	38.92	42.43	38.91	42.43	-3.51	- 3.52	0.01
90th	49.0	53.14	48.41	52.38	-4.14	-3.97	0.17
Mean	23.74	34.16	24.346	34.06	-10.42	-9.71	0.71

[a] Single mother, SM for brevity.
[b] Single woman without children, SW for brevity.

without children in pre-WFTC. Table 2 shows exactly this. There are statistically significant differences between single mothers and single women without children before the introduction of the WFTC. In pre-WFTC, a single woman without children worked around 34 hours/week, while a single mother worked nearly 24 hours/week. There is a 10.42 hour/week difference which is significant at the 1 percent level. The share of women working more than 16 and 30 hours/week provides a better picture of the labor market behaviors. On average, 58.6 percent of single women without children were working more than 16 hours/week, compared to only 27.3 percent of single mothers - less than half the share of single women without children. On the other hand, nearly 49 percent of single women without children were working more than 30 hours/week in my sample, compared to only 16 percent of single mothers - a reduction by almost three times.

Table 2: Variables by treatment group, pre-WFTC

	Single Woman	Single Mother	Diff (SW - SM)	Diff. missing values
Paid hours based on actual hours	34.164	23.744	10.420***	-0.230***
Working 16+ hours	0.586	0.273	0.313***	-
Working 30+ hours	0.485	0.161	0.324***	-
Age	35.806	25.451	10.355***	-
Educational qualification			-0.652***	-0.0176***
- Degree or equivalent	18.57	4.58		
- Higher education	9.01	4.63		
- GCE A level or equivalent	20.97	16.93		
- GCE grade A-C or equivalent	22.86	38.57		
- No qualification	15.60	22.05		
- Other	12.19	12.73		
Ethnicity			-0.132***	0.0254***
- White	94.42	91.27		
- Black	3.13	5.24		
- Indian/Pakistani/Bangladeshi	1.42	2.24		
- Other	0.192	0.21		
Accommodation			-0.265***	0.000256***
- Owned property	20.44	8.94		
- Being bought with mortgage or loan	42.61	45.39		
- Rented	35.45	44.53		
- Rent-free	1.08	0.67		
- Other	0.07	0.10		
Region of place of work			0.409***	-0.184***
Religion			-0.593***	-
- Catholic	71.03	66.69		
- Presbyterian	4.99	4.42		
- Church of Ireland	4.18	4.01		
- Other	19.58	24.88		
Observations	230,465	180,015		

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4.3 Independent Variables

The independent variables are chosen to capture the background characteristics that explain the labor market participation of my sample. The literature agrees that age, educational qualification, ethnicity, accommodation type, region of place of work, and religion are the main determinants of labor hours and labor force participation (Blundell et al., 2005; Leigh, 2005; Gregg and Harkness, 2003; Brewer et al., 2005; Francesconi and van der Klaauw, 2004; Dickens and McKnight, 2008; Blundell and Hoynes, 2001). Table A1 shows the summary statistics for the independent variables. My sample consists of single women of working age (15 - 64 years old), where the average age is 25.5 years old. Table 2 shows that single mothers tend to be approximately 10 years younger than single women without children, where the average age of single women without children (single mothers) is 34 (24) years old. Educational qualification describes the type of education the woman in my sample had, and contains 7 categories: degree or equivalent (1), higher education (2), GCE A-level or equivalent (3), GCE grade A-C or equivalent (4), other qualification (5), no qualification (6), and "don't know" (7). The GCE is a General Certificate of Education, which could be standard or advanced (A-level). Table 2 shows that nearly 18 percent of single women without children have a degree or an equivalent, compared to only 4.6 percent of single mothers. 23 (38) percent of single women without children (single mothers) have a GCE with grades A through C, and 21 (17) percent obtained a GCE A-level. Single mothers are less educated than single women without children, with that difference being statistically significant. Ethnicity contains 9 categories: white (1), black-caribbean (2), black-african (3), black- not caribbean or african (4), indian (5), pakistani (6), bangladeshi (7), chinese (8), other(9). Single mothers and single women appear to be of the same ethnic background, although single mothers are less likely to be White and are more likely to be Black and Indian/Pakistani/Bangladeshi, with the difference being significant. However, over 90 percent of both single women without children and single mothers are White. Religion also contains 9 categories: catholic (1), presbyterian (2), church of Ireland (3), methodist (4), other protestant (5), other religion (6), no denomination (7), does not apply (8) and unwilling to answer (9). 71 (66) percent of single women without children (single mothers)are catholic. Accommodation contains 7 categories: owned property (1), being bough with mortgage or loan (2), part rent part mortgage (3), rented (4), rent-free (5), squatting (6), non-contact (7). Single mothers are twice as less likely to own property compared to single women, and are almost 10 percent more likely to be renting their current accommodation. There are statistically significant differences in the region the women work at, although the regions are omitted from the table for brevity.

The characteristics of single mothers in my sample are consistent with the WFTC literature. On average, single mothers are younger, less educated, less likely to own property and more likely to be renting their houses, and are more likely to be non-White. This implies that my sample is representative of single women in the UK during the 1990s. There are statistically

significant differences in all independent variables; therefore, I will control for them in my empirical specification. The missing values for paid hours, educational qualification, ethnicity, accommodation, and region are statistically non-random (table 2, last column). To minimize the selection problem, I control for all missing values in my analysis.

5 The Quantile Difference-In-Difference Approach

5.1 Distributional Treatment Effects

Nonexperimental program evaluation methods based on multiple pre- and post-treatment periods are conducted in different ways. The literature on in-work benefits in the UK, US, and Europe generally estimates the Average Treatment Effect (ATE) or Local Average Treatment Effects (LATE) through discontinuity designs[13] to estimate the causal effect of in-work benefits on labor force participation (Bettendorf et al., 2014). Although most research in program evaluation focuses on estimating the average effect of a policy or program, in some cases it is also useful to discover the effect at different locations of the distribution. It is often of interest to know whether a policy made individuals at the lower tail of the distribution better off (e.g., the poor are on the lower tail of the income distribution, and we want to know whether a policy has a positive effect on the poor). If a policy affects people at different ends of the distribution in the same way, then this policy has homogeneous treatment effects and going beyond the ATE is not insightful. However, when the policy creates effects of different magnitudes and directions at different locations of the distribution, there are heterogeneous treatment effects (Heckman et al., 1997). These effects are often referred to as distributional treatment effects, where Firpo (2007) and Callaway and Li (2017) identify such effects under the presence of selection on observables, under an instrumental variable (Frölich and Melly, 2013; Abadie et al., 2002), or when accessing repeated cross-sections data (Athey and Imbens, 2006).

In this paper, I apply the quantile difference-in-difference identification strategy (QDID) to recover the quantile treatment effects (QTE) of the WFTC on single mothers' paid labor hours. Quantile regressions provide an indication of the magnitude of the impact, while being much less sensitive to extreme values than the mean. The DID assumptions which are generally applied to recover the ATE (Card and Robins, 1996) can only partially identify the QTE because the quantile function is not a linear operator like the expectation. Several papers give an overview of this methodology and its extensions (Angrist and Krueger, 1994; Meyer et al., 1995; Arkhangelsky, 2019; Blundell and MaCurdy, 1999). However, despite the prevalence of the DID method, there has been little empirical work in the last three decades that examine the assumptions for the identification of the QTE under QDID, especially when having access to repeated cross-section data. Callaway and Li (2017) discuss the assumptions for the identification of the QTE specifically when accessing panel data, Athey and Imbens (2006) examine the assumptions for the identification of the QTE under the QDiD framework when accessing repeated cross-sections data, while several papers discuss applications of this methodology (Meyer et al., 1995; Finkelstein and McKnight, 2008; Pomeranz, 2015; Havnes and Mogstad, 2015).

[13] Difference-In-Difference (DID), Regression Discontinuity Design (RDD), and Difference-In-Difference-In-Difference (DDD)

5.2 The Quantile DID Approach

I follow the non-linear DID identification strategy (Athey and Imbens, 2006) to analyze the effect of the WFTC at each quantile of the distribution of paid labor hours. Athey and Imbens (2006) propose a new model, named the Changes-In-Changes model (CIC), which is a generalization of the standard DID model (Angrist and Krueger, 1994). The CIC differs from the standard DID in three aspects: first, the CIC allows for heterogeneous time and treatment effects across individuals; second, they propose an estimator for the untreated (control) group to recover the Average Treatment on the Untreated (ATU); and third, they allow for selection on treatment effects. The evaluation problem in observational studies is that we can only observe individuals in one state (treated or untreated) and it is difficult to find valid counterfactuals for the treated group. Valid counterfactuals ensure that the estimator does not pick up any other differences non-related to the policy intervention. Athey and Imbens (2006) also develop the QDID approach, in which they compute the counterfactual distribution of the outcome variable by adding the change over time at the qth quantile of the control group to the qth quantile to the first-period treatment group. I use this approach in this thesis, which has also been applied to Meyer et al. (1995) and Poterba et al. (1995).

The standard DID model is as follows. Suppose individual i belongs in one of the groups $D_i \in \{0,1\}$ (where 1 is the treatment group) and is observed in one of the time periods $P_i \in \{0,1\}$. Let Y_i^1 denote the outcome for the treated, and Y_i^0 the outcome for the control. According to Rubin's Causal Model (Rubin, 1974), treatment T_i is an indicator $T_i = D_i * P_i$, which equals to 1 when individual i belongs to $D_i = 1$ and is in the post-treatment period $P_i = 1$. Therefore, the observed outcome for each individual i is

$$Y_i = T_i Y_i^1 + (1 - T_i) Y_i^0$$

which is equal to Y_i^1 when $T_i = 1$ and equal to Y_i^0 when $T_i = 0$.

In the standard DID model, the structural equation for the control is

$$Y_i^0 = \alpha + \beta P_i + \gamma D_i + \theta X_i + \varepsilon_i^0 \tag{1}$$

and the structural equation for the treated group is

$$Y_i^1 = \alpha + \beta P_i + \gamma D_i + \theta X_i + + \delta_i T_i + \varepsilon_i^1 \tag{2}$$

In each equation, coefficient β represents the time effect, γ represents group-specific effects, and θ represents the effects of the exogenous independent variables X_i. The last term, $\varepsilon_i^{T_i}$, $\forall T_i \in \{0,1\}$, represents the individual-specific unobservable term. This term is assumed to be exogenous from the time and group-specific effects and assumed to have the same distribution

over time

$$\varepsilon_i^{T_i} \perp (P_i, D_i, T_i)$$

Therefore, the standard DID estimator is

$$\begin{aligned}\delta_i^{DID} = \big[\mathbb{E}[Y_i|P_i=1,D_i=1] - \mathbb{E}[Y_i|P_i=0,D_i=1]\big] - \\ \big[\mathbb{E}[Y_i|P_i=1,D_i=0] - E[Y_i|P_i=0,D_i=0]\big]\end{aligned} \tag{3}$$

The standard DID model assumes that the treatment effects are homogeneous (e.g.: $\delta_i = \delta$, $\forall i$) and that the mean impact is sufficient to understand the effects of a policy. Under the DID assumptions, we can consistently estimate the coefficients using OLS and obtain $\delta^{DID} = \delta^{ATE}$. The ATE is informative about the mean impact of the entire distribution of Y_i. However, there are many cases where the researcher suspects heterogeneous treatment effects and wants to estimate the Treatment on the Treatment at different quantiles of the distribution of Y_i. This estimator is named the Quantile Treatment Effect and can be estimated under the QDID approach. Athey and Imbens (2006) discuss the identification of this estimator, which relies on predicting an outcome for the counterfactual group and subtracting it from the outcome of the treated group.

Assume that the conditional distributions of Y_i^0 and Y_i^1 are F_{Y^0} and F_{Y^1} with supports $\mathbb{Y}_{dp}^0$ and $\mathbb{Y}_{dp}^1$, respectively. For $q \in [0,1]$ and for a random variable Y with support $\mathbb{Y}$, we have

$$F_y^{-1} = inf\{y \in \mathbb{Y} : F_Y(y) \geq q\} \tag{4}$$

This implies that the inverse distribution functions are continuous from the left and that for all $q \in [0,1]$ we have $F_Y(F_Y^{-1}) \geq q$ for any continuous Y. Using the below transformation, we denote $k(y)$ as the second-period outcome for an individual in the control group, $D_1 = 0$.

$$k(y) = F_{Y_{0,1}}^{-1}(F_{Y_{0,0}}(y)) \tag{5}$$

Meyer et al. (1995) assume that the median of Y^0 conditional on D_i and P_i is equal to $\alpha + \beta P_i + \gamma D_i$. Given that the median is also a quantile, we can apply this approach to each quantile q and obtain coefficients $\alpha_q, \beta_q, \gamma_q$ for the control group. Therefore,

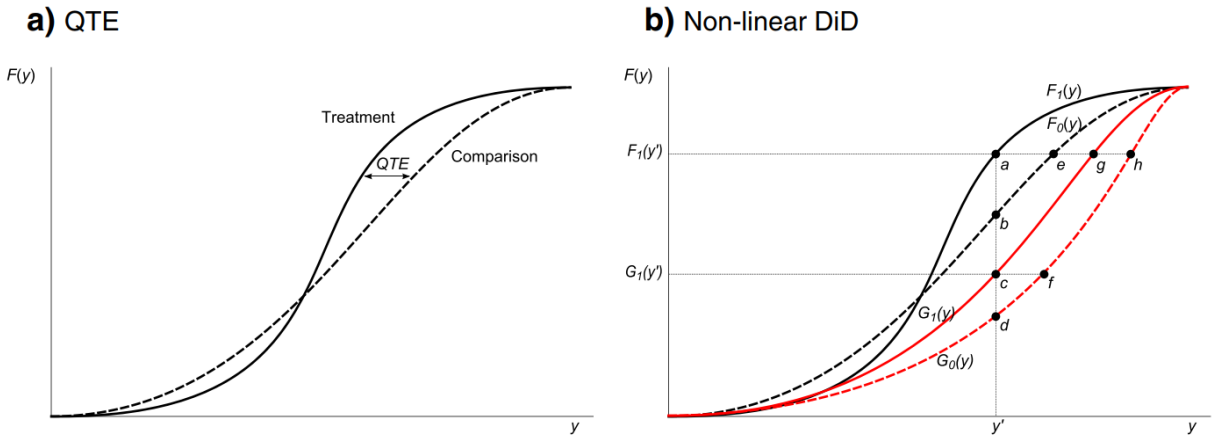
$$k(y)^{QDID} = F_{Y_{0,1}}^{-1}(F_{Y_{0,0}}(y)) = F_{Y_{1,0}}(q) + \Delta^{QDID} \tag{6}$$

where

$$\Delta^{QDID} = F_{Y_{01}}^{-1}(q) - F_{Y_{00}}^{-1}(q)$$

is the difference over time for the control group at a given quantile q. This difference is added to $F_{Y_{1,0}}(q)$, which is the distribution of the treatment group in pre-treatment, to obtain the predicted counterfactual value $k^{QDID}(y)$. With this approach, I compare individuals

Figure 6: Illustration of the empirical strategy: Quantile treatment effects and non-linear DID estimates. Source: Havnes and Mogstad (2015)



Note: $F_t(y)$ and $G_t(y)$ are the distributions for the treated and control group at t=0 and t=1. The CIC estimate: (a-e) - (c-f), the QDD estimate: (a-e) - (g-h).

across both groups and time according to their quantile. Under continuity of the outcome Y, the predicted counterfactual of the standard DID model and of the QDID are identical: $\mathbb{E}[k^{DID}(Y_{10})] = \mathbb{E}[k^{QDID}(Y_{10})]$, becoming easier to estimate[14]. Figure 6, panel A shows how the QTE is estimated by taking the horizontal differences between the two distributions, while panel B shows how the CIC and QDID procedure take place.

The identification of the QTE under the QDID framework requires four assumptions.

Assumption 1

$$Y^0 = h(\varepsilon, P, T)$$

implying that the outcome of an individual in absence of the treatment should be a function of time and group-specific effects, and unobservables.

Assumption 2 The production function $h(\varepsilon, P, T)$ is strictly increasing in ε

Assumption 3 The following condition must hold $\forall q \in [0,1]$

$$\varepsilon \perp (D, P)$$

Assumption 4 The distribution of unobservables must be identical in all subpopulations (quantiles). This eliminates within-quantile heterogeneity.

It is important to note that the parallel trends assumption, a typical assumption in the standard DID model, is not necessary for the estimation of the QTE under the CIC model. Athey and Imbens (2006) loosen many of the assumptions of the standard DID model and that is why they argue that the standard DID model is nested in their CIC model. In the standard DID

[14]For more technical details and proofs, please refer to Athey and Imbens (2006)

model, groups and time periods are treated symmetrically: the mean of the individual outcomes in the absence of the treatment is additive in group and time fixed-effects. This is the classical "parallel trends" assumption. In contrast, the CIC model allows for time periods and groups to be treated asymmetrically in pre-treatment, and even accommodates selection on treatment benefits (the Ashenfelter's dip). On the other hand, the QDID approach requires the parallel trends assumption, making it more restrictive.

5.3 Identification Strategy

My identification strategy is the following: I compare paid labor hours before and after the introduction of the WFTC between eligible women (single mothers) and non-eligible women (single women without children). The periods before the WFTC cover Spring 1996 - Winter 1998, and the periods after the WFTC cover Summer 2000 - Winter 2002. The treatment group consists of all single mothers, and the control group is single women without children. My sample consists of all women who were working in pre-WFTC. Keeping only incumbents avoids the ambiguity of whether the estimates are driven by incumbents or those women that entered the labor market post-WFTC. The evidence in the literature of the effect of the WFTC on labor hours is mixed because the studies include working and non-working single women. Additionally, keeping only incumbents makes it easier to assume that single mothers remain in the same rank pre- and post-WFTC[15]. Therefore, my QTE estimates are more easily interpreted when keeping incumbents only.

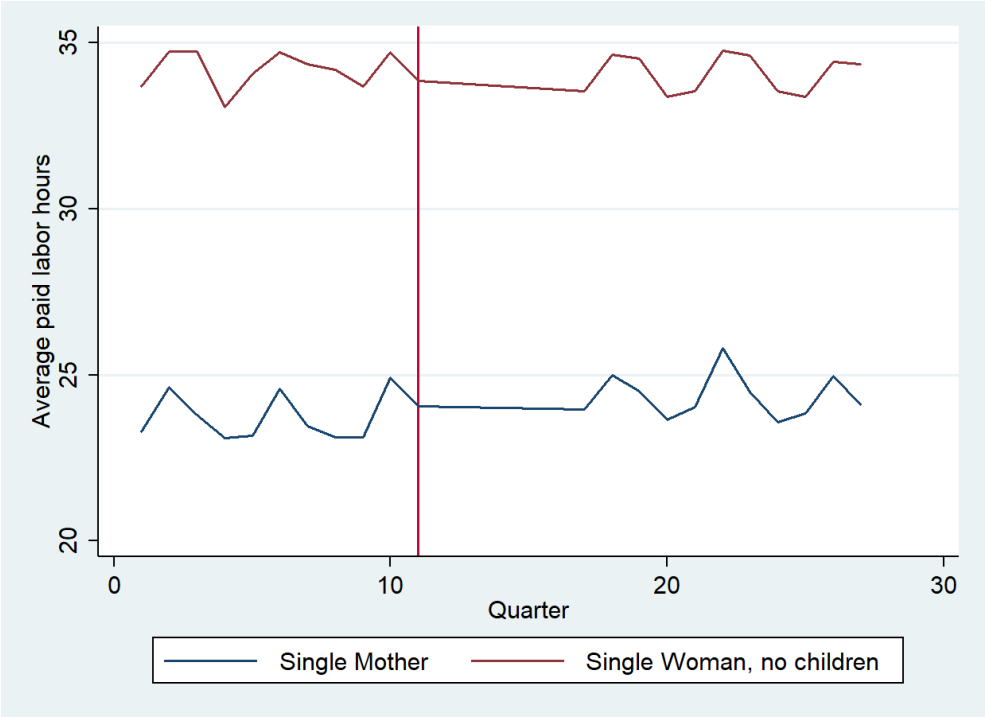
Figure 7 shows the trends of paid labor hours for the treatment and control group. The trends move in an almost-parallel pattern, implying that in pre-WFTC the unobservables and observables also move in the same pattern. My identification strategy raises two key concerns: (i) the treatment and control group are non-random because a woman self-selects into becoming a single mother or a single woman without children; therefore, there is presence of selection; and (ii) the parallel trends assumption may not be satisfied when formally testing for it. The selection problem occurs when individuals select into treatment or control based on observable characteristics (independent variables) and/or on unobservables. I address issue (i) by conditioning my analysis on the independent variables. The conditional QTE makes identification easier because it eliminates selection on observables. Conditioning on covariates is the closest observational studies can do to replicate experimental studies (Heckman et al., 1997). On the other hand, the DID approach assumes that

$$\mathbb{E}[u_{ip_1} - u_{ip_0} | D_i = 1] = \mathbb{E}[u_{ip_1} - u_{ip_0} | D_i = 0]$$

where u_{ip_1} and u_{ip_0} are unobservables in period P_1 and P_2.This condition is normally satisfied when the parallel trends assumption is satisfied. I address issue (ii) in my sensitivity

[15]Rank invariance in quantile regressions assumes that the individuals remain in the same quantile pre- and post-intervention.

Figure 7: Paid hours trends pre- and post-WFTC, by group



analysis in section 6.

5.4 Empirical Specification

An intervention can cause heterogeneous changes for the treated individuals at different locations of the distribution of the outcome variable. However, there is no formal way of making sure that heterogeneous treatment effects exist before performing a full analysis. In general, there should be some indication that looking at the mean impact of an intervention on the outcome variable could disguise significant heterogeneity. I choose to estimate the QTE because I suspect heterogeneity regarding paid labor hours for two reasons: first, figure 8 conveys the effect of moving the distribution of paid labor hours from pre- to post-WFTC. The movements are consistent with table 1. It shows that, unconditional on other labor-market-related variables, the WFTC positively increased the paid labor hours for the lower quantiles, while it had no effect for the top quantiles. Although this figure does not statistically *prove* that there are between-quantile heterogeneous effects, it gives *sufficient* evidence that the movements of the CDF are not homogeneous and that there is reason to suspect the existence of treatment heterogeneity. Second, the variance of paid hours is large relative to the mean (see table A1). A large variance of the data implies large dispersion around the lower or higher ends of the distribution. This suggests that looking at the mean impact of the WFTC on paid hours would conceal significant heterogeneity that cannot be captured unless using quantile treatment effects.

In this paper, the primary outcome variable is the number of paid hours a woman works in

the week she was interviewed. At any qth quantile, the following relationship holds:

$$Y_{it} = \alpha + \beta Post99_{it} + \gamma SingleMother_{it} + \theta X_{it} + \delta T_{it} + \varepsilon_{it} \tag{7}$$

or

$$Q(Y_{it}|Post99_{it}, SingleMother_{it}, X_{it}, T_{it}) = \alpha_q + \beta_q Post99_{it} + \gamma_q SingleMother_{it} + \theta_q X_{it} + \delta_q T_{it} \tag{8}$$

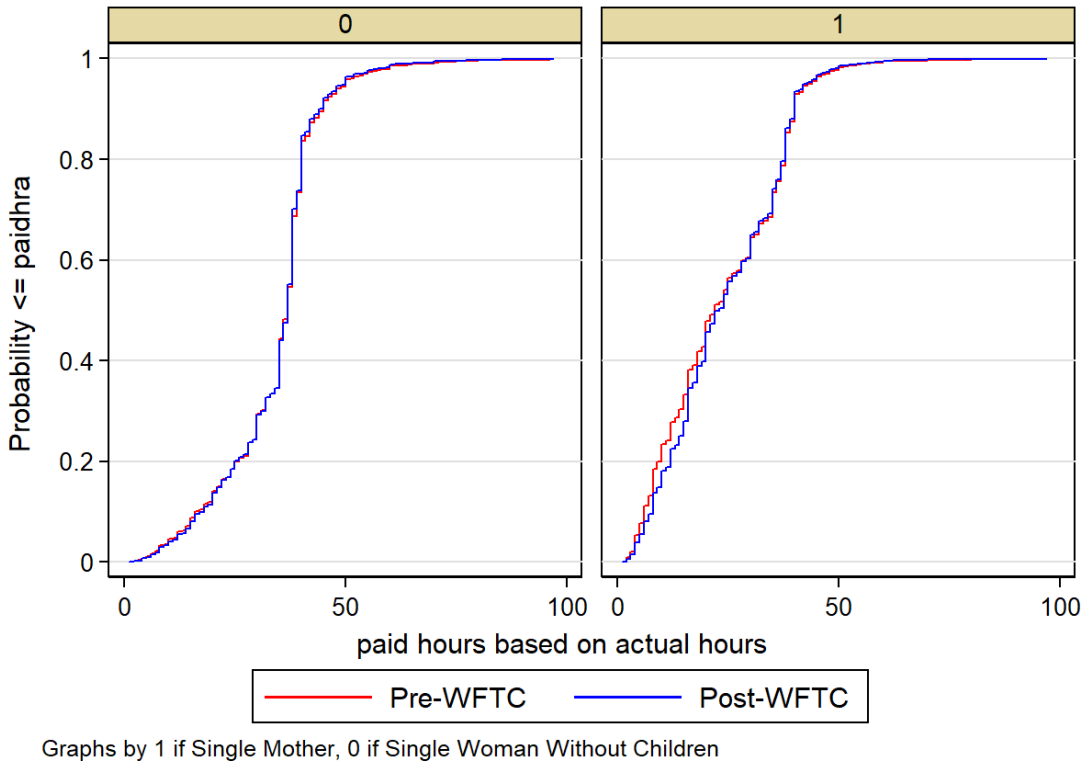
where Y_{it} is the observed number of paid labor hours for each woman; $Post99 = 1\{Quarter \geq 17\}$, which corresponds to the immediate quarter after Spring 2000, when the WFTC was fully phased in; $SingleMother = 1\{Children > 0\}$; T = Post99 * SingleMother, equals to 1 if the single mother is observed after the introduction of the WFTC; X_{it} (for the conditional QTE) are all the independent variables, including age, age squared (experience), educational qualification, ethnicity, accommodation, region of place of work, and religion. I also add dummy variables for each variable that has missing values, which takes the value 1 if that value is missing in the corresponding variable.

The coefficient of interest is δ_q. This coefficient estimates the effect of WFTC on single mothers at each specified quantile. For ease of interpretation, δ_q estimates the following difference at each quantile q.

$$\begin{aligned}\delta_q = &\Big[\text{ Paid Hours(Post-99, Single Mother)} - \text{Paid Hours(Pre-99,Single Mother)}\Big] - \\ &\Big[\text{Paid Hours(Post-99, Single Woman)} - \text{Paid Hours(Pre-99, Single Woman)}\Big]\end{aligned} \tag{9}$$

To interpret the assumptions required for the QDID model, I first assume that the structural model of labor supply for the control group, Y^0, increasing in the unobservables ε. I interpret ε as the number of hours a woman would desire to work if that woman was not eligible for the WFTC (i.e., does not have children). It seems plausible to assume that the number of paid hours a single woman would work is increasing in her desire to work in the absence of the WFTC. Additionally, Assumption 3 seems valid because the distribution of ε should stay the same over time within the same group D_i. This implies that the changes in the WFTC do not lead to rapid and drastic changes in labor supply decisions, which seems plausible.

Figure 8: The CDF of Paid Hours for single women without children (to the left) and single mothers (to the right)



6 Results

6.1 The Quantile Treatment Effects

In this section, I present my results from applying the QDID approach as suggested by Athey and Imbens (2006). Given that my choice for estimating the QTE was based on figure 8, I first replicate this figure and estimate the following equation:

$$Y_{it} = \alpha_q + \beta_q Post99_{it} + \gamma_q SingleMother_{it} + \delta_q T_{it} + \varepsilon_{it} \tag{10}$$

This would yield the unconditional QTE. However, the unobservables ε_{it} potentially contain other labor-market-related variables that are endogenous to the decision of the paid labor hours. Specifically, age, age-squared (experience), educational qualification, ethnicity, accommodation, region of place of work, and religion are characteristics that explain the labor market behavior of any individual. Moreover, being a "single mother" is not exogenous to labor market decisions, and may very well be correlated to the mother's background characteristics. However, there may be worry that the effect of the WFTC could be mediated through other variables, including experience, accommodation, and education. It could be that the effect WFTC was higher for those with more working experience, or that mothers may have acquired more

education because of the WFTC. Additionally, accommodation was affected by the WFTC because the WFTC was counted as income when calculating housing benefits. Because of the ambiguity of the mediation, I estimate two conditional QTE estimates. The first conditional QTE - *basic conditional* - contains all background information that could not be affected by the WFTC: age, religion, region, and ethnicity. The second conditional QTE - *standard conditional* - controls for all independent variables and their missing values: age, age squared, educational qualification, ethnicity, accommodation, region, religion.

Table 3 shows the main results of this thesis. For each specification, I provide the ATE and the QTE at each quantile. For ease of interpretation, next to each quantile I show the number of hours the single mothers were working in pre-WFTC.

Unconditional QDID

The unconditional QTE suggests that on average, the WFTC increased the single mothers' labor hours by 0.7 hours/week. The ATE, however, disguises significant treatment heterogeneity. For mothers in the bottom quantiles, the WFTC significantly increased labor hours between 1-3 hours/week, which corresponds to an increase by 4.5 - 45 percent. For mothers in the top quantiles, there is no statistically significant effect on labor hours. The unconditional QTE provides evidence similar to figure 8 and table 6 and suggests that for the bottom quantiles the substitution effects dominates the income effects, leading to an overall increase in the single mothers' labor hours.

Basic conditional QDID

The basic conditional QTE draws a different picture than the unconditional QTE. The effect remains the same for single mothers in the bottom quantiles, where the largest increase is in the 20th quantile. Table 3 suggests that single mothers working 8.5 hours/week pre-WFTC increased their labor hours by more than 6.6 hours, making them eligible for the WFTC. On the other hand, for single mothers in the top quantiles, the WFTC significantly reduced labor hours between 1-5 hours/week. The largest reduction was for mothers working around 30 hours/week pre-WFTC (60th quantile). Although the WFTC payments increased again once the mother worked 30 hours or more per week, it seems that these mothers had a higher preference for non-labor activities, and that the increase in the credit was not high enough to induce them into working full-time. The ATE suggests a small but insignificant increase, suggesting that the substitution effect offset the income effect.

Standard conditional QDID

The standard conditional QTE shows similar evidence to the basic conditional QTE. The effect is weaker for the bottom quantiles and there is a stronger negative effect in the top quantiles relative to the basic conditional QDID. This implies that the effect of the WFTC might have been mediated through education, experience, or accommodation. The bottom quantiles show

a softer increase compared to the unconditional or basic conditional QTE. It seems that the only significant increase is for mothers who were working 8.5 hours/week pre-WFTC (20th quantile). The income effect in the top quantiles is stronger in this specification, with the strongest reduction for the 60th and 70th quantile (both around the 30 hour/week threshold). Education, experience, and/or accommodation are responsible for the large negative effect in the top quantiles. I explore this relationship in section 7. The ATE suggests that the income effect dominates the substitution effect when I control for labor-market-related variables.

Figure 9 plots the estimates for each specification. Overall, the coefficients suggests that the WFTC significantly increased labor hours for mothers between the 10th and 30th quantile. For the bottom quantiles, the substitution effect dominates the income effect. When controlling for other labor-market-related variables, the WFTC significantly reduced paid labor hours for single mothers between the 60th to 80th quantile, or those mothers working between 29.5 - 39 hours/week in pre-WFTC. Thus, for the top quantiles, the income effect dominates the substitution effect. Although the estimates vary across specification, there are some consistent patterns. The 20th quantile consistently exhibits the largest increase, while labor hours drop the most on the 60th quantile. The effect on single mothers around the median (40th 50th quantile) is overall weakly significant and small. This suggests that the WFTC was not very effective for single mothers working between 16 - 30 hours/week. However, it is clear that the WFTC was very effective to induce single mothers into part-time employment (16 hours/week), but not into full-time employment (30 hours/week).

Table 3: Quantile and Average Treatment Effects

	Outcome: Paid hours based on actual work		
	Unconditional QDID	**Conditional QDID[a]**	**Conditional QDID[b]**
DiD (ATE)	0.701***	0.121	-1.520**
	(0.000)	(0.821)	(0.030)
Quantile (Pre-WFTC hours/week)			
10th (4.5 hours/week)	2.099***	3.189***	0.923
	(0.000)	(0.000)	(0.150)
20th (8.5 hours/week)	3.099***	6.683***	2.118***
	(0.000)	(0.000)	(0.002)
30th (14 hours/week)	2.099***	1.501**	0.756
	(0.000)	(0.014)	(0.447)
40th (19 hours/week)	2.099***	0.115	-0.432
	(0.000)	(0.914)	(0.651)
50th (23 hours/week)	1.099**	-0.508	-2.603*
	(0.018)	(0.723)	(0.067)
60th (29.5 hours/week)	0.0990	-4.836***	-5.438***
	(0.882)	(0.000)	(0.000)
70th (35.5 hours/week)	0.0990**	-2.696***	-4.206***
	(0.041)	(0.000)	(0.000)
80th (39 hours/week)	0.0990	-0.725*	-3.279***
	(0.683)	(0.071)	(0.000)
90th (49 hours/week)	0.0990**	0.353	-2.722***
	(0.041)	(0.416)	(0.000)
N	227948	24012	23992

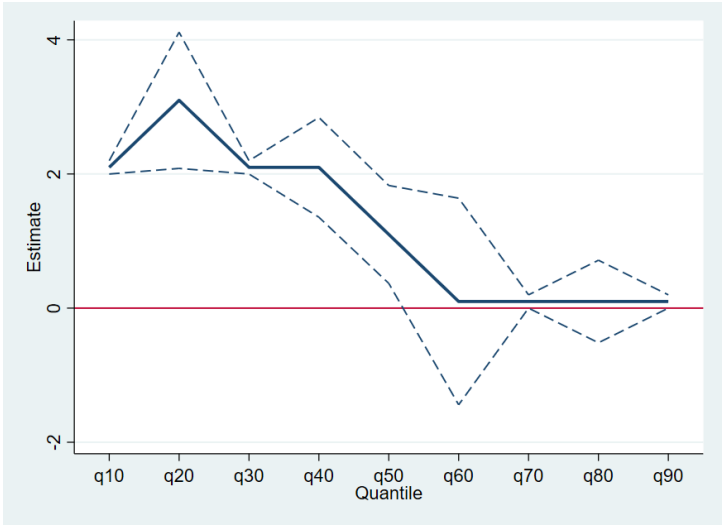
p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

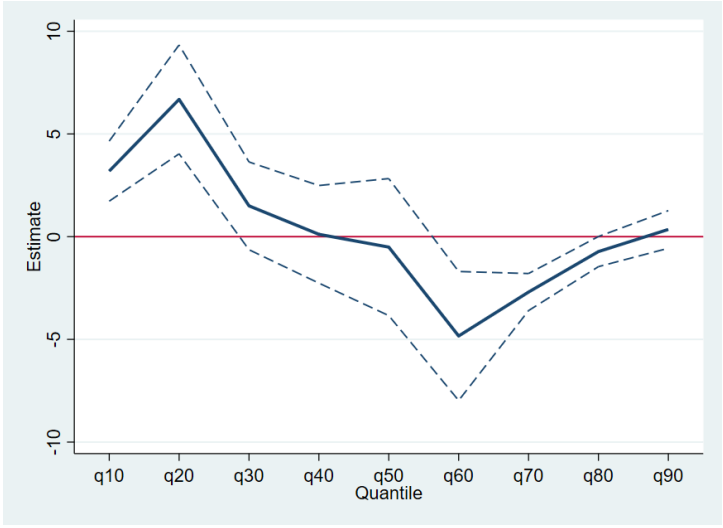
[a] Controlling for age, region, religion, and ethnicity

[b] Controlling for age, age squared, region, religion, ethnicity, accommodation, educational qualification, and for the missing values

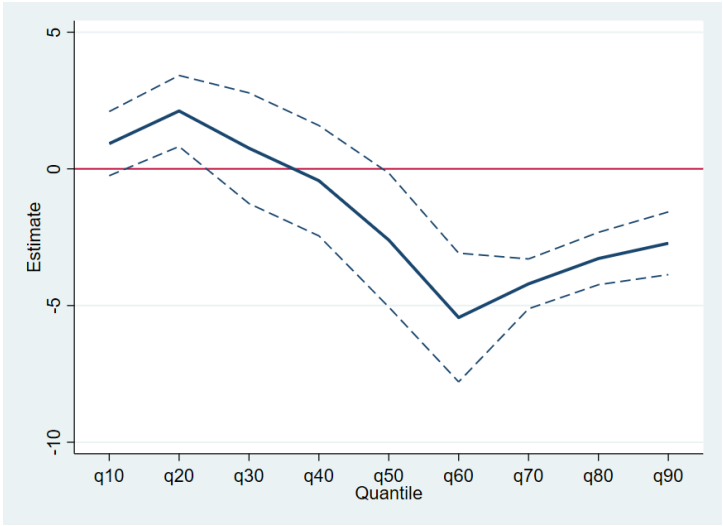
Figure 9: QTE estimates by specification



(a) Unconditional QTE



(b) Basic Conditional QTE



(c) Standard Conditional QTE

6.2 Characteristics by Quantile

My QTE estimates suggest that there indeed is treatment heterogeneity regarding the hours worked in pre-WFTC. Microeconomic theory suggests that the workers' preference for labor and leisure determines how the worker responds to an increase in labor income. Given that the QLFS does not have data on individual preferences, in this subsection I describe the single mothers' characteristics by quantile to undercover the reasons why the substitution effect dominates the income effect in the bottom quantiles, and why the income effect dominates the substitution effect for the top quantiles.

Figure A.10a shows the average hourly pay by quantile[16]; figure A.10b shows the average age of the youngest child by quantile; and figure A.10c shows the average age of the single mothers by quantile. The figures suggest that single mothers in the:

> **Bottom quantiles** tend to be younger, get paid less by the hour, and have "older" younger children
>
> **Top quantiles** tend to be older, get paid more by the hours, and have "younger" younger children

These trends are purely descriptive and do not imply a causal effect; nevertheless, it may offer some insight that explain the substitution and income effects[17]. Single mothers in the bottom quantiles had a lower hourly pay in pre-WFTC which could also explain why they were working less hours. The WFTC increased the returns to working and the incentive to work, and as a result leisure became more expensive to single mothers. It is possible that the mothers saw the WFTC as an opportunity to work more hours and have higher income. Additionally, mothers in the bottom quantiles were younger, but also had "older" younger children relative to the top quantiles. This suggests that younger mothers were more productive and having older children allowed them to work more.

The WFTC did not encourage single mothers in the top quantiles to work more. Those mothers had a higher hourly pay in pre-WFTC. An increase in income as a result of the WFTC increased their demand for all normal goods, which also included leisure. The mothers' preferences shifted towards leisure, resulting in a reduction of labor hours. Additionally, those mothers were older and had "younger" younger children. It is generally acknowledged that older workers have a higher preference for leisure and are less productive at work. Moreover, the presence of "younger" younger children may have encouraged them to work less so they could dedicate their leisure hours to spending time with their children.

[16] These are the same quantiles as in table 3. They divide the CDF of paid labor hours of single mothers pre-WFTC into 9 quantiles.

[17] I do not perform a heterogeneity analysis for independent variables. The STATA command that I use in my thesis, *cic.ado*, does not show the effect of the variables at each quantile, but only for the mean.

6.3 Did the WFTC Encourage Full-time Employment?

Figure A.11 shows the evolution of the distribution of hours of work among single mothers pre- and post-WFTC. The kernel density estimates are closely related to histograms, but it is advantageous due to its smoothness and continuity of the data. In pre-WFTC, there were a bunching of data around 8.5 hours/week (the 20th quantile), 16 hours/week (part-time employment) and 30 hours/week (full-time employment). This is probably due to the rules under the FC, which lowered the minimum requirement of part-time employment from 24 hours/week to 16 hours/week and set full-time employment at 30 hours/week. This figure suggests that the WFTC significantly lowered the share of single mothers working 8.5 hours/week and increased the share of single mothers working 16 hours/week. The share of mothers working full-time weakly increased post-WFTC. A pattern consistent with previous evidence occurs for the 20th quantile, suggesting that mothers working 8.5 hours/week pre-WFTC significantly increased their paid labor hours to become eligible for the WFTC. Therefore, there is strong evidence that the WFTC pushed single mothers into part-time employment, but there is overall *weak* evidence that it pushed them into full-time employment.

6.4 Robustness Checks

As always in policy evaluation with observational data, researchers cannot guarantee that the estimates do not suffer from selection bias, but a series of sensitivity analysis can provide enough proof for the credibility of the findings. It is important that the effect of the WFTC is similar across various specifications at each quantile rather than just the mean. In this subsection, I run three sensitivity analysis: (i) I check the identification assumption; (ii) I run falsification tests using alternative definitions of the treatment and time variable, and (iii) I check whether my estimates hold under the CIC model (Athey and Imbens, 2006).

Identification assumption

The identifying assumption of the QDID estimator is that the change in population shares from before to after treatment around a given level of paid labor hours (quantile) would be the same in the treatment and control group in the absence of treatment. A concern is that the QTE estimates pick up differential time trends or the effects of any omitted variables. There are two alternatives to test for this concern: (i) by adding controls to the empirical estimation to eliminate concerns for omitted variable bias, and (ii) to test for the parallel trends assumption. Columns 2 and 3 in table 3 perform the first alternative. They suggest that the QTE estimates differ across the empirical specifications and that the unconditional QTE of the WFTC may have picked up the effect of omitted variables, especially for the top quantiles. For the second option, I first run a regression where I interact the treatment indicator with each quarter and test for the joint significance of the first 10 interaction terms (pre-WFTC quarters). The F-test suggests that the pre-WFTC interaction terms between the treatment and time periods are statistically

significant, which implies that my QTE estimates may be picking up differential time trends. The failure of the parallel trends assumption is consistent in the WFTC literature, although the main outcome variable was employment. Figure 7 shows the graphical representation of the trends of paid labor hours of work for single mothers and single women without children pre- and post-WFTC. Although the trends look similar, there is a slight increase for single mothers in pre-WFTC relative to single women without children. Gregg and Harkness (2003) argue that this increase would have occurred *irrespective* of the introduction of the WFTC because of the policies that were introduced in 1998 (NDLP) which specifically targeted single parents. Blundell et al. (2005) and Leigh (2005), on the other hand, argue that the increase was temporary and would not have continued after 1998. They argue that the spurious growth in employment was due to the introduction of many policies that overlapped in 1998 specifically targeting low-income or single-parent households. Francesconi and van der Klaauw (2004; 2007) are the only studies that attribute the convergence in 1998 to the "anticipation effect". Nevertheless, this is questionable. There was no campaign or publicity regarding the WFTC in 1999 because it was only replacing an existing policy (Brewer and Browne, 2006). Additionally, it would seem insensible for single mothers to adjust their labour hours *one year prior* to ensure that they receive the WFTC allowance in October 1999. Therefore, the literature agrees that the changes in the employment rate (and labour hours) pre-WFTC occurred independently of the WFTC.

Falsification checks using alternative Treatment and Time definition

The WFTC literature contains a large discussion on the choice of the control group for single mothers. There have been multiple concerns (Bettendorf et al., 2014; Francesconi and van der Klaauw, 2004; Eissa and Liebman, 1996) that single women without children are not a valid control group because they are more educated, implying that their unobservables may also differ. Naturally, single mothers may resemble low-educated single women without children more. Looking at their summary statistics (not shown), low-educated single women are indeed more similar to single mothers: they work less hours pre-WFTC (30 hours/week versus 34 hours/week for all single women), 30 percent work 16 hours or more per week (versus 58.6 percent of all single women), and only 18 percent work 30 or more hours (versus 48.5 percent of all single women). Table A2 shows the QTE estimates when restricting the control group to low-educated single women without children. The estimates are slightly smaller, but the pattern seems to be consistent to table 3. This gives more credibility my original estimates and the choice of the control group.

Alternatively, I test for differential time trends by redefining the post-treatment period. I define a pseudo-post-treatment period, *post97*, if the woman was observed after the last quarter in 1997. I repeat the same estimation procedure, and if my estimates suggest highly significant treatment effects, then it would suggest that my initial QTE pick up differential time trends.

Table A3 shows the results for the pseudo-post-treatment period. The results for the unconditional QTE are highly similar to the original results in table 3; however, the results for the conditional QTE are *weakly significant* and do not show the same patterns (there is no drop in labor hours in the 60th quantile). It would be desirable to perform this test for other periods, for instance, when the WFTC was actually announced (in early 1998)[18]. Given that other policies were introduced in late 1998, I could essentially be picking up the effects of the NDLP. This is an issue in the WFTC literature, and I cannot confidently say that my QTE estimates show the effect of the WFTC *only*. Most of the literature agrees that their estimates identify the combined effect of each policy that was introduced in the late 1990s (Blundell et al., 2005; Leigh, 2005; Blundell and Hoynes, 2001; Francesconi and van der Klaauw, 2004; Francesconi and Van der Klaauw, 2007; Gregg and Harkness, 2003; Dickens and McKnight, 2008; Brewer, 2001). The violation of the parallel trends assumption and the weak significance of the QTE under a pseudo-post-treatment period suggests that my QTE estimates may pick up the effect of previous policies that increased the single mothers' labor supply.

Alternative Non-linear DID Estimates

To evaluate the sensitivity of the QTE estimates to the assumption about the counterfactual of the paid hours distribution and to the parallel trends assumption, I also implement the CIC estimator (Athey and Imbens, 2006). The CIC model uses the observed changes in the control distribution to estimate the counterfactual distribution of the treatment group in the post-treatment periods. The CIC estimator is a three-step procedure (Havnes and Mogstad, 2015). For every value of y in the support of $F_{1,1}(y)$, first find the quantile of y in the control distribution $F_{0,1}(y)$; second, find the paid hours level at that quantille in the control distribution in pre-treatment, $F_{0,0}(y)$; and third, find the quantile of that paid hours level for the treatment distribution $F_{1,0}(y)$. The CIC estimator for the counterfactual distribution is

$$k(y)^{CIC} = F_{1,0}(F_{0,0}^{-1}(F_{0,1}(y))) \tag{11}$$

The advantage of the CIC model is that it does not require that groups and time periods are treated symmetrically in the pre-treatment periods, unlike the QDID. The parallel trends assumption is not required for the estimation of the CIC estimator because the definition of the time period and group takes a different notation (please refer to (Athey and Imbens, 2006) for a more detailed explanation). In the CIC model, groups can differ in arbitrary ways in the distribution of the unobserved individual characteristics, and allows for selection on treatment (implying that it also allows for those mothers who adjusted their labor hours pre-WFTC to benefit from the WFTC). Therefore, if the CIC estimates are similar to the QDID estimates, it is reassuring that the assumptions on the counterfactual distribution are not causing bias.

[18]Defining the post-treatment period earlier than winter 1997 resulted in multicollinearity and the estimations could not be performed.

Table A4 shows the results from the CIC model. The estimates change slightly from the QDID model. For the unconditional CIC, the mean and the quantile effects are similar, with no effects between the 60th and 90th quantiles. For the basic and standard conditional CIC, the estimates are less significant but follow the same patterns and magnitudes. The mean impact is almost identical to the QDID model and the effects are in the same direction. The conditional CIC estimates suggest that the substitution effect dominates the income effect for the bottom quantiles, while the income effect dominates the substitution effect for the top quantiles. The similarity between the CIC and QDID estimates suggest that the QDID estimates are not biased due to the assumptions I made on the counterfactual distribution. Nevertheless, the CIC model does not work so efficiently under the presence of covariates (Melly and Santangelo, 2015), which is why I do not choose the CIC model for my main analysis.

7 Discussion

Under different specifications, my mean impacts are consistent with previous evidence: the unconditional QDID yields a significantly positive mean impact (Gregg and Harkness, 2003), insignificant effect under the basic conditional QDID (Francesconi and Van der Klaauw, 2007), and significant negative effect under the standard conditional QDID (Meyer and Rosenbaum, 2001). This signals how inconsistent mean impacts are when analyzing the effect of in-work benefits on labor hours. On the other hand, the QDID yields more consistent estimates from which I can observe a pattern: the WFTC increased labor hours for single mothers in the bottom quantiles because of the substitution effect, and it decreased labor hours for single mothers in the top quantiles because of the income effect. My estimates are robust regarding the definition of the treatment and control group; however, falsification checks find that my estimates pick up differential time trends from other policies that were introduced in the late 1990s. Nevertheless, given the similarity between my QTE estimates under the QDID and CIC model, it is reassuring that assumptions on the distributions are well-defined.

My results are consistent with a labor supply model where mothers in the lower quantiles of labor hours work more because of an increase in labor income, while mothers in the higher quantiles work less because they prefer leisure over labor hours. The reasons behind these mechanisms are ambiguous. In section 6.1 I found that the effect of the WFTC was mediated through education, experience, and/or accommodation. Accommodation in particular has been discussed extensively in the literature because of the interaction between the WFTC and housing benefits (HB). In my dataset, accommodation captures the housing arrangements of single mothers and single women without children. Given that single mothers were renting more (table 2), this variable may capture the effect of HB on labor supply behavior. The QLFS does not provide complete information regarding HB[19]; therefore, in this section, I discuss possible explanations of the adverse labor supply responses.

The WFTC payments were calculated based on the number of hours worked, the number and age of children, and weekly net income. Out-work income (e.g., HB, child benefit) was disregarded from net income; however, the WFTC payments were counted as income when calculating the other benefits, specifically HB. During the economic growth in the late 1990s, rent grew rapidly, pushing the UK government to raise the HB by the same rate. HB was one of the fastest-growing elements of the social benefits system (Giles et al., 1997), with the government spending more than £5 billion per year. Nevertheless, the interaction of the WFTC and HB pushed many researchers to worry about the financial incentives to work for the poorer parents, leading to potential unemployment and poverty traps.

[19]Some variables captured who was on HB and the amount of HB that they received, but it was not recorded in all quarters and there were many missing values.

The basic formula for the HB payment was

$$\text{HB} = \text{Rent} - (\text{Taper} * \text{Excess income}) - \text{Non-dependent deduction} \tag{12}$$

where excess income is

$$\text{Excess income} = (\text{Earnings} - \text{Disregard}) + \text{Other income} - \text{Needs} \tag{13}$$

Other income included other social security benefits, including child benefits and tax credits. The taper in equation 12 is 65 percent, implying that an increase in excess income by £1 reduced the HB to only 35 pence. Figure 3 illustrates the interaction of the WFTC and HB. While on income support (yellow area in figure 3b) the entire rent was paid by the HB. But when single mothers worked more than 16 hours/week, the HB was reduced by 65 percent. When the single mother worked 30 hours/week, the burden of the rent was entirely upon her. Therefore, she faced two options: (i) work more hours so she could afford rent; however, the extra hours she could afford to work may not have been enough to cover rent; or (ii) work less and benefit more from the HB.

When the WFTC was introduced, many policymakers voiced their concerns that the WFTC would create distorted labor market effects, with strong incentives to work 16 hours/week, but little incentive to work full-time. Figure 3b shows the net income once the parent became eligible for the WFTC. The incentive to work 16 hours/week under the WFTC is clear, which is shown by the high jump in net income once the parent works 16 hours/week. However, the withdrawal rate created little incentive for single parents to work more hours as they were approaching 30 hours/week. Dilnot and McCrae (1999) predicted that the greatest gains from the WFTC would be for parents in the lower tail of the income distribution. Assuming that paid hours is increasing in income, my findings suggest that the bottom quantiles of the paid hours distribution benefitted more from the WFTC. Their behaviour is consistent with the substitution effect, where single mothers started working more to become eligible for the WFTC. The WFTC made *part-time* employment more attractive only, and working 16 hours/week led to an appreciable increase in income.

The effect for the top quantiles is more complicated. The marginal tax rate (MTR) indicates the change in net income from a small change in the working hours. When MTR is high, there is very little incentive to work. The WFTC was calculated to have a 70 percent MTR for single mothers, compared to 97 percent under the FC (Dilnot and McCrae, 1999). Among all eligibles for the WFTC, single mothers faced the lowest incentive to work under the WFTC (Dilnot and McCrae, 1999). The negative effect on work incentives of high marginal withdrawal rates extending up the income distribution as taper rates were cut has often been ignored in the literature. Where high MTRs exist over a wide range of income, increasing hours of work may yield very little extra income, and reducing hours would result in little loss of net income but

a large increase in leisure time (Giles et al., 1997). Therefore, the withdrawal rate when the single mother worked between 16 - 30 hours/week and the interaction of HB with WFTC could potentially explain why single mothers in the top distribution of paid hours (assuming that they earn more money) significantly drop their labor hours for leisure. This behaviour would be consistent with the income effect.

8 Conclusion

In this thesis, I investigate the effect of the WFTC on single mothers' paid labor hours. I apply the Quantile Difference-In-Difference approach (QDID) and estimate the Quantile Treatment Effects (QTE). This approach allows me to investigate the heterogeneous treatment effects of the WFTC given the initial hours of work the single mother was working pre-WFTC. I use the UK Quarterly Labor Force Survey and obtain a sample of more than 411,000 working single women, with and without children, where single women without children are my control group. I obtain data from Spring 1996 to Winter 2002 and drop the data from 1999 and the first quarter in 2000, for a total of 22 quarters. I keep only single women who were employed in pre-WFTC in my sample for three reasons: first, to eliminate the effect of those entering the labor market post-WFTC; second, to make rank invariance more plausible; and third, to provide clear evidence of the effect of the WFTC on labor hours because the findings from the literature are mixed and ambiguous.

I estimate the unconditional and conditional QTE. Both specifications suggest that the WFTC significantly increased labor hours for single mothers in the bottom quantiles (working between 4.5 - 14 hours/week in pre-WFTC) by 2- 6 hours/week. The highest increase was consistently for the 20th quantile (8.5 hours/week). For mothers in this group, the substitution effect dominates the income effect, resulting in an increase in labor hours. The effect of the WFTC for the top quantiles differs by the specification. The unconditional QTE finds no effect, while the conditional QTE suggests that the WFTC lowered paid labor hours for the top quantiles by 1-5 hours/week. For single mothers in this group, I argue that the income effect dominates the substitution effect, encouraging mothers to prefer and substitute leisure over labor. The QTE estimates are highly robust to the choice of the control group and the assumptions made on the distribution of the counterfactual group, but are less robust to the differential time trends. Therefore, I cannot confidently say that my QTE estimates do not contain the effects of other policies that were introduced shortly before the WFTC.

The findings of this thesis provide handy evidence of the effect of the WFTC on single mothers' labor market decisions. I contribute to the WFTC literature in two ways: first, this is the first study to analyze treatment heterogeneity of the WFTC regarding paid labor hours; and secondly, I find evidence that the WFTC shifted single mothers to part-time employment, but did not shift single mothers to full-time employment. There might be two possible reasons why single mothers in the top quantiles did not work more because of the WFTC. First, the WFTC taper rate may have discouraged single mothers from working more because the marginal credit did not increase net income considerably. Second, the interaction between WFTC and housing benefits may have discouraged mothers from working because the housing benefits were decreasing in WFTC. A greater effect might have been achieved if the UK government increased the reward at the starting level (16 hours/week) rather than spreading the payments across the labor hours distribution. My findings, however, should be taken with caution because the re-

sults are sensitive to the empirical specification and to differential time trends. Nevertheless, the WFTC unequivocally provided strong financial incentives for parents to work part-time. These results can provide fresh insights to policymakers inside or outside of the UK who aim to design family tax policies to stimulate labor market participation.

9 References

Abadie, A., J. Angrist, and G. Imbens (2002). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica 70*, 91–117.

Angrist, J. and A. B. Krueger (1994). Why do world war ii veterans earn more than nonveterans? *Journal of Labor Economics 12*, 74–97.

Arkhangelsky, D. (2019, 03). Dealing with a technological bias: The difference-in-difference approach.

Athey, S. and G. W. Imbens (2006, 03). Identification and inference in nonlinear difference-in-differences models. *Econometrica 74*, 431–497.

Bettendorf, L. J. H., K. Folmer, and E. L. W. Jongen (2014, 11). The dog that did not bark: The eitc for single mothers in the netherlands. *Journal of Public Economics 119*, 49–60.

Bingley, P. and I. Walker (1997, 09). The labour supply, unemployment and participation of lone mothers in in-work transfer programmes. *The Economic Journal 107*, 1375–1390.

Blundell, R., M. Brewer, and A. Shephard (2005, 06). Evaluating the labour market impact of working families' tax credit using difference-in-differences.

Blundell, R. and H. Hoynes (2001). Has "in-work" benefit reform helped the labour market?

Blundell, R. and T. MaCurdy (1999). Labor supply: A review of alternative approches.

Borjas, G. J. (2020). *Labor economics*. New York, Ny Mcgraw-Hill Education.

Brewer, M. (2001). Comparing in-work benefits and the reward to work for families with children in the us and the uk. *Fiscal Studies 22*, 41–77.

Brewer, M. and J. Browne (2006). The effect of the working families' tax credit on labour market participation.

Brewer, M., A. Duncan, A. Shephard, and M. J. Suarez (2005, 06). Did working families' tax credit work? the final evaluation of the impact of in-work support on parents' labour supply and take-up behaviour in the uk.

Brewer, M. and P. Gregg (2003, 05). *Eradicating child poverty in Britain: welfare reform and children since 1997*. Policy Press.

Callaway, B. and T. Li (2017). Quantile treatment effects in difference in differences models with panel data. *SSRN Electronic Journal*.

Card, D., C. Michalopoulos, and P. Robins (1999). Measuring "wage progression" among former welfare recipients.

Card, D. and P. K. Robins (1996, 03). Do financial incentives encourage welfare recipients to work? early findings from the canadian self sufficiency project.

Dickens, R. and D. T. Ellwood (2001, 04). Whither poverty in great britain and the united states? the determinants of changing poverty and whether work will work.

Dickens, R. and A. Manning (2004). Has the national minimum wage reduced uk wage inequality? *Journal of the Royal Statistical Society. Series A (Statistics in Society) 167*, 613–626.

Dickens, R. and A. McKnight (2008, 10). The impact of policy change on job retention and advancement.

Dilnot, A., J. Kay, and C. Morris (1984). The reform of social security.

Dilnot, A. and J. McCrae (1999). Family credit and the working family tax credit.

Eissa, N. and H. Hoynes (2004). Taxes and the labor market participation of married couples: the earned income tax credit. *Journal of Public Economics 88*, 1931–1958.

Eissa, N. and J. B. Liebman (1996, 05). Labor supply response to the earned income tax credit. *The Quarterly Journal of Economics 111*, 605–637.

Ellwood, D. T. (1999, 11). The impact of the earned income tax credit and social policy reforms on work, marriage, and living arrangements.

Finkelstein, A. and R. McKnight (2008, 07). What did medicare do? the initial impact of medicare on mortality and out of pocket medical spending. *Journal of Public Economics 92*, 1644–1668.

Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica 75*, 259–276.

Francesconi, M. and W. van der Klaauw (2004). The consequences of "in-work" benefit reform in britain: new evidence from panel data.

Francesconi, M. and W. Van der Klaauw (2007). The socioeconomic consequences of "in-work" benefit reform for british lone mothers.

Frölich, M. and B. Melly (2013). Unconditional quantile treatment effects under endogeneity. *Journal of Business Economic Statistics 31*, 346–357.

Giles, C., P. Johnso, and J. McCrae (1997). Housing benefit and financial returns to employment for tenants in the social sector. *Fiscal Studies 18*, 49–72.

Gregg, P. and S. Harkness (2003). Welfare reform and lone parents employment in the uk.

Hansard (1997, 07). London: The stationery office.

Havnes, T. and M. Mogstad (2015, 07). Is universal child care leveling the playing field? *Journal of Public Economics 127*, 100–114.

Heckman, J. J. (1993). What has been learned about labor supply in the past twenty years? *The American Economic Review 83*, 116–121.

Heckman, J. J., J. Smith, and N. Clements (1997, 10). Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *The Review of Economic Studies 64*, 487–535.

Hoffman, S. D. and L. S. Seidman (2003). *Helping Working Families: The Earned Income Tax Credit*. W.E. Upjohn Institute for Employment Research.

Hotz, V., C. Mullin, and J. Scholz (2006). Examining the effect of the earned income tax credit on the labor market participation of families on welfare.

Internal Revenue Service (1996). Earned income credit, for use in preparing 1996 returns.

Keane, M. and R. Moffitt (1998, 08). A structural model of multiple welfare program participation and labor supply. *International Economic Review 39*, 553.

Leigh, A. (2005). Optimal design of earned income tax credits: Evidence from a british natural experiment. *Center for Economic Policy Research*.

Meghir, C., D. Phillips, S. Adam, R. Blundell, M. Brewer, E. French, and G. Laroque (2009). and taxes.

Melly, B. and G. Santangelo (2015). The changes-in-changes model with covariates.

Meyer, B. D. and D. T. Rosenbaum (2001, 08). Welfare, the earned income tax credit, and the labor supply of single mothers*. *The Quarterly Journal of Economics 116*, 1063–1114.

Meyer, B. D., W. K. Viscusi, and D. L. Durbin (1995). Workers' compensation and injury duration: Evidence from a natural experiment. *The American Economic Review 85*, 322–340.

OECD (1996). Earnings inequality and mobility.

Office for National Statistics (2002). The quarterly labour force [data set].

Office for National Statistics (2015). User guide volume 3 - details of lfc variables.

Paull, G., I. Walker, and Y. Zhu (2000, 03). Child support reform: Some analysis of the 1999 white paper. *Fiscal Studies 21*, 105–140.

Pomeranz, D. (2015, 08). No taxation without information: Deterrence and self-enforcement in the value added tax. *American Economic Review 105*, 2539–2569.

Poterba, J. M., S. F. Venti, and D. A. Wise (1995, 09). Do 401(k) contributions crowd out other personal saving? *Journal of Public Economics 58*, 1–32.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology 66*, 688–701.

Sanctuary, C. J. and K. R. Nurse (1981). Evaluation of changes to family income supplement. *Omega 9*, 469–480.

Tyger, A. (2019, 07). Earned income tax credit reform.

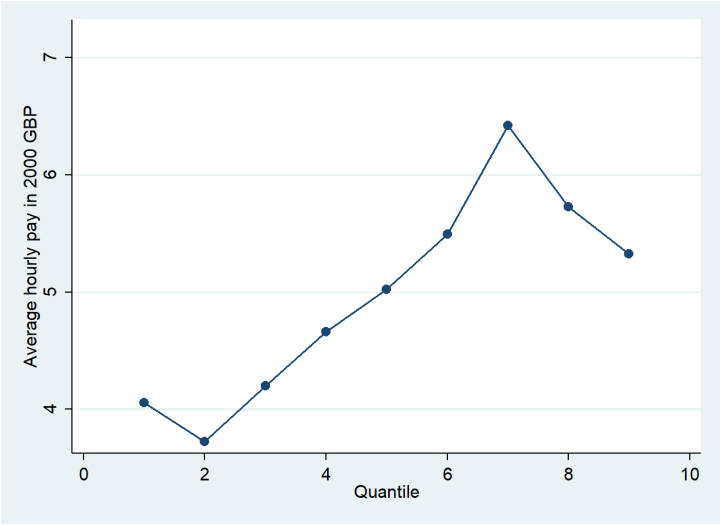
UK Data Archive (n.d.). User guide lfs two-quarter and five-quarter longitudinal datasets.

A Appendix

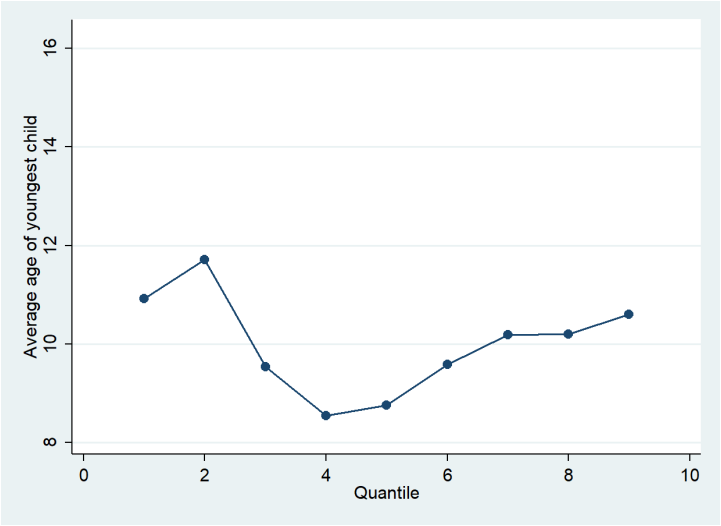
Table A1: Summary statistics, pre-WFTC

	Mean	S.D	Min	Max	N
Paid hours	23.74	13.82	1	97	113,634
Working 16+ hours	.27	.44	0	1	209,659
Working 30+ hours	.16	.36	0	1	209,659
Age	25.45	9.98	15	64	209,659
Educational qualification	4.26	1.31	1	7	182,795
Ethnicity	1.34	1.37	1	9	166,396
Accommodation	2.83	1.13	1	7	209,505
Region	12.12	5.53	1	22	71,971
Religion	2.92	2.44	1	9	7,467

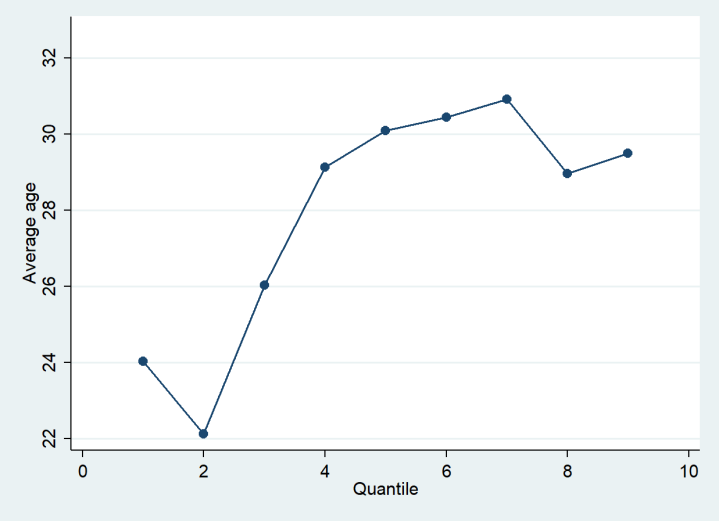
Figure A.10: Characteristics of single mothers in pre-WFTC by quantile



(a) Hourly pay by quantile



(b) Age of youngest child by quantile



(c) Age of single mother by quantile

Table A2: Control: Low-educated single women without children

	Outcome: Paid hours based on actual work		
	Unconditional QDID	**Conditional QDID[a]**	**Conditional QDID[b]**
DID (ATE)	0.810***	-0.604	-1.659**
	(0.000)	(0.462)	(0.036)
Quantile			
10th	0.350***	2.338***	0.188
	(0.000)	(0.009)	(0.817)
20th	2.350***	5.346***	2.219***
	(0.000)	(0.002)	(0.009)
30th	2.350***	2.084	0.750
	(0.000)	(0.187)	(0.543)
40th	1.350***	-0.862	-1.206
	(0.000)	(0.567)	(0.290)
50th	0.350	-4.002*	-3.601**
	(0.338)	(0.054)	(0.013)
60th	1.350**	-5.712***	-3.386***
	(0.033)	(0.000)	(0.001)
70th	-0.650	-2.917***	-3.800***
	(0.365)	(0.000)	(0.000)
80th	0.350	-1.263**	-3.375***
	(0.348)	(0.030)	(0.000)
90th	0.350***	-0.196	-2.059**
	(0.000)	(0.717)	(0.011)
N	164837	14812	14804

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[a] Controlling for age, region, religion, and ethnicity

[b] Controlling for age, age squared, region, religion, ethnicity, accommodation, educational qualification, and for the missing values

Table A3: Alternative time definition: Post97≥ Spring 1998

	Outcome: Paid hours based on actual work		
	Unconditional QDID	**Conditional QDID[a]**	**Conditional QDID[b]**
DID (ATE)	0.474***	0.896	-0.690
	(0.001)	(0.391)	(0.498)
Quantile			
10th	1.069***	2.485***	0.917
	(0.000)	(0.002)	(0.295)
20th	2.069*	6.418**	2.528**
	(0.070)	(0.010)	(0.029)
30th	2.069***	2.204	1.698
	(0.000)	(0.207)	(0.198)
40th	1.069**	0.461	0.779
	(0.028)	(0.800)	(0.626)
50th	0.0692	0.631	-0.195
	(0.741)	(0.668)	(0.929)
60th	0.0692	-0.0542	-2.396
	(0.935)	(0.984)	(0.274)
70th	0.0692	-2.099	-3.250*
	(0.325)	(0.148)	(0.075)
80th	0.0692	-0.873	-3.038***
	(0.325)	(0.139)	(0.001)
90th	0.0692	-0.197	-2.330**
	(0.325)	(0.723)	(0.020)
N	227948	24012	23992

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[a] Controlling for age, region, religion, and ethnicity

[b] Controlling for age, age squared, region, religion, ethnicity, accommodation, educational qualification, and for the missing values

Table A4: CIC estimates

	Outcome: Paid hours based on actual work		
	Unconditional CIC	**Conditional CIC[a]**	**Conditional CIC[b]**
CIC			
Mean	0.422***	0.310	-1.409*
	(0.000)	(0.630)	(0.087)
Quantiles			
10th	1**	3.938***	2.968***
	(0.047)	(0.009)	(0.006)
20th	2***	5.834***	2.439***
	(0.000)	(0.000)	(0.003)
30th	1***	0.656	0.719
	(0.000)	(0.586)	(0.521)
40th	2***	-0.111	0.309
	(0.000)	(0.937)	(0.766)
50th	1**	0.173	-1.354
	(0.034)	(0.889)	(0.362)
60th	0	-1.526	-4.969**
	(1.000)	(0.443)	(0.012)
70th	0	-2.131***	-3.887***
	(1.000)	(0.000)	(0.000)
80th	0	-0.830***	-4.028***
	(1.000)	(0.000)	(0.000)
90th	0	-0.363	-4.165***
	(1.000)	(0.355)	(0.000)
N	227948	24012	23992

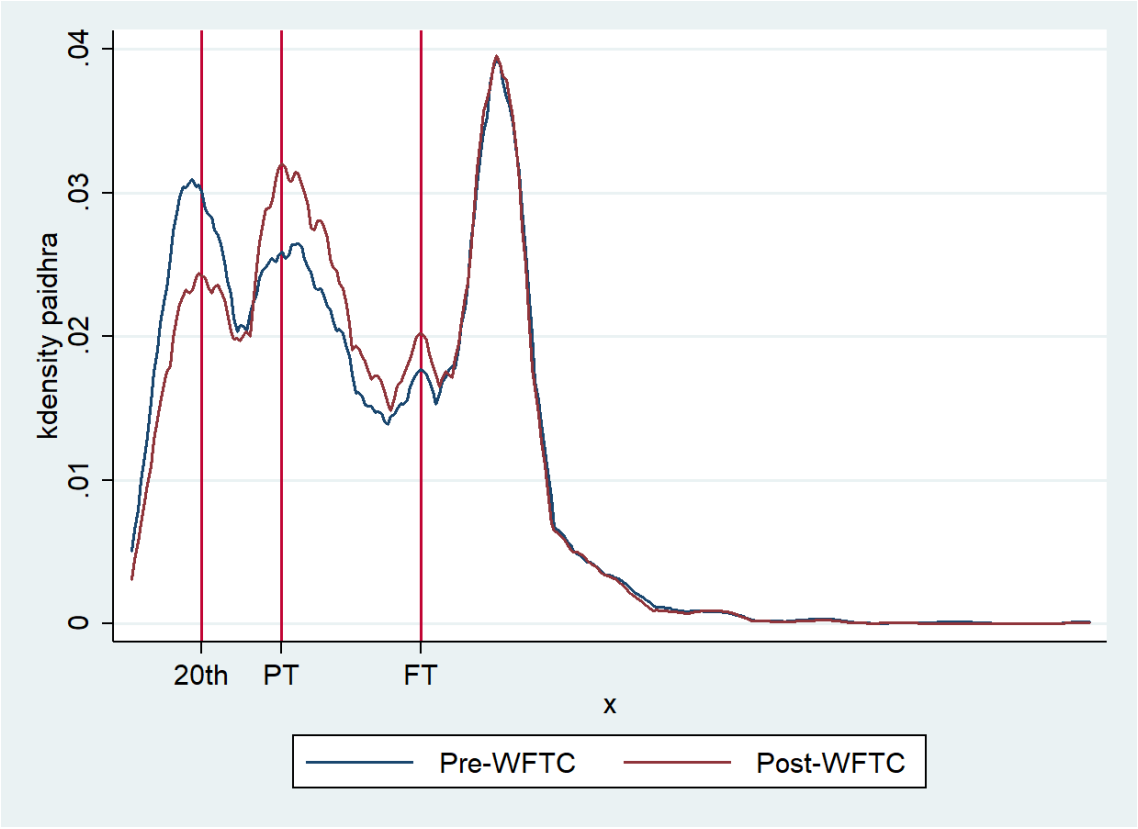
p-values in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

[a] Controlling for age, region, religion, and ethnicity

[b] Controlling for age, age squared, region, religion, ethnicity, accommodation, educational qualification, and for the missing values

Figure A.11: Kernel Density Distribution of Paid Hours pre- and post-WFTC



List of Figures

List of Tables