Oxford Economic Papers Advance Access published December 18, 2015

Oxford Economic Papers, 2015, 1–21
doi: 10.1093/oep/gpv068

British tax credit simplification, the intra-household distribution of income and family consumption

By Paul Fisher

Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ; e-mail: pfishe@essex.ac.uk

Abstract

This paper asks whether targeting welfare benefits to women can be effective at changing household spending. We provide empirical evidence on this question by using a reform to the UK tax-credit system in 2003 as a quasi-experiment. We find that the reform caused low-income households to reallocate spending towards children's goods. The results further demonstrate that the effects of directing welfare benefits to women can extend beyond child expenditures to goods that are collectively consumed by all household members. Our findings are in contrast to those from earlier studies that took place in the economic setting of 1970s UK.

JEL classifications: H31, I38, J16

1. Introduction

This paper provides new evidence on whether paying welfare benefits to women in couples (instead of men) can change household spending patterns. This is important, as evidence that households respond differently to transfer payments, depending on the gender of the recipient, has implications for the design of the programs. In general, it is difficult to assess whether the income of the male or female partner in a couple has a differential impact on household spending as income shares are endogenous. We address this issue by exploiting a novel reform to the UK tax-credit system in 2003 as a quasi-experiment. The reform made the 'carer of the children' the recipient of a large share of a household's tax-credit income, which is in contrast to the pre-reform rules where the male partner could generally claim ownership of all means-tested child payments and tax-credits.

This paper seeks to answer these three questions: i) Do families spend transfer income differently depending on the gender of the recipient, ii) Do children benefit when welfare benefits are paid to mothers (instead of fathers) in terms of increased household spending on children's goods, and at the cost of adult expenditures, iii) Do any effects extend to collectively consumed goods? We use variation in the share of female benefit income in total household benefit income generated by the 2003 UK tax-credit reform. The reform effect is estimated for multiple spending items using a difference-in-differences estimator that exploits the fact that low-income households are more likely to be affected by the reform, relative to middle income households. Gregg *et al.* (2006) use a similar strategy when studying the introduction of a UK in-work benefit—the Working Families Tax Credit, in 1999—on household spending. However, their results do not inform us of expenditure patterns had WFTC been paid directly to mothers instead of fathers.

The paper relates to previous research that considered redirecting transfer income associated with a UK child benefit reform (replacing tax expenditures received by men with transfers that went to women), which took place in the setting of the 1970s (Lundberg *et al.*, 1997; LPW; Ward-Batts, 2008; WB; Hotchkiss, 2005). LPW find evidence that households shifted expenditure away from male clothing and towards female and children's clothing, in line with the conventional claim that women attach more weight to children's (and their own) welfare. An absence of randomized control trials from advanced economies in contemporary time periods has hampered testing, whether such a relationship holds in the setting of a modern developed economy.[1]

Changes to the economic and social setting since the 1970s, including increased female employment rates and a smaller gender wage gap, mean that it is not possible to apply the LPW results to the setting of this paper, that is, at the start of the 21st century. Thus our main contribution is that our findings apply to an advanced economy in a recent time period and so inform our understanding of the household decision making process as societies develop. Separately, much of the earlier research focused on a narrow group of adult and children's goods for which consumption can be assigned to a particular household member (e.g., male, female, and children's clothing). But women are also likely to face different incentives to men to invest in other publicly-consumed household goods. For example, gains from marriage such as housing investments may be captured by one person if the marriage breaks down (Stevenson, 2007).[2] As such, the effects of directing benefit income to women on household spending potentially extend beyond a narrow range of women and children's goods that have received most attention in the literature. In light of the above, we extend our analysis to collectively consumed household expenditures.

In terms of the main findings, we conclude that the 2003 reform influenced spending patterns through the intra-household distribution of income. We observe spending increases on children's goods (toys and games, musical instruments) and spending decreases on private adult goods (gambling, maintenance payments). However, the strongest effects are found on items that are consumed by all household members. Insofar as these goods are consumed by children as well as adults, the overall effect of directing state benefits to women on children's consumption is unclear. Finally, we contrast our results to the LPW and WB studies of the child benefit reforms that took place three decades ago. Explanations are put forward for the key similarities and differences.

1 A closely-related pool of causal evidence is available in the development literature. Braido *et al.* (2012) find evidence of income pooling in Brazil. In contrast, studies in various settings have shown that children benefit when the bargaining position of women is improved; see Duflo (2000) (South Africa); Duflo and Udry (2004) (Cote d'Ivoire); Bobonis (2009) (Mexico); and Attanasio and Lechene (2010) (Mexico).

2 Stevenson (2007) finds that laws regarding the division of marital property at divorce affects newlyweds' home ownership decisions.

The paper proceeds as follows. The next section provides some background on the 2003 UK tax-credit reform before our empirical strategy is discussed in Section 3, followed by a description of the data in Section 4. In Section 5, we provide some background to the 1970s LPW reforms, and contrast the 1970s samples to those in this paper. The empirical results are presented in Section 6, while Section 7 concludes. A summary of results from robustness checks is included in Appendix 1. Full tables of robustness checks are available in the supple mentary material.

2. Background

A reform that aimed to simplify the UK tax-credit system in 2003 provides the opportunity to test whether redirecting state benefits from husbands to wives influences household spending patterns in the setting of the 21st century. Tax-credits are in-work benefits that top-up the income of poor families meeting certain employment conditions. Prior to 2003, the Working Families Tax Credit (WFTC) operated in the UK and was typically paid with male wages. In 2003, WFTC was split into two new tax credits: the Working Tax Credit (WTC) and Child Tax Credit (CTC). In couples, while WTC was again usually paid with male wages, CTC was redirected as a cash payment to the partner designated the ‘carer of the children’ (usually the mother).

Tax-credit eligibility requires that at least one partner in a couple has paid employment for at least 16 hours per week. Prior to 2003, tax credits were normally paid in the wage packet of the partner satisfying this weekly hours requirement, unless couples requested that the payment be made to the non-working partner. This implies three categories of tax-credit eligible couples defined by the partner satisfying the weekly work hours condition: the male, both partners, or the female.[3] For the first group of couples, payment was typically made through the wages of the male partner. The second group of couples chose which partner received payment, and so the mother could only receive payment with the male’s consent. The final, smaller group of couples are atypical in that the benefit system gave the female partner ownership of tax credits. To give an indication of the size of this final group, for the treated group in our data they constitute 9.4%.

In April 2003, the new tax credits came into operation. The key point from the perspective of this paper is that payments of the new tax credits were divided between the partners. In this way, the reform increases the ability of mothers in couples to lay claim to a substantial share of tax-credit income. One could argue that couples have a degree of choice over the partner designated as the ‘main carer’ since the couple indicates who this person is when completing the tax-credit application form. However, the available tax-credit statistics by gender report that in October 2003, 87% of CTC payments in couples went to the female partner (Inland Revenue, 2003, table 7.1).[4]

Figure 1 illustrates the relevant features of the reform for couples with different gross earnings. The figure takes a single-earner couple with two children and the male working 30 hours per week. It then varies gross weekly earnings (essentially varying the male wage rate) to show how the value of tax credit entitlement changes. For a couple with set gross weekly earnings, the impact of the reform on total tax credit income can be seen by

3 Self-employed workers were paid directly and are excluded from the estimation sample.

4 In addition to WFTC, families paying income tax could claim a small tax rebate known as the Children’s Tax-Credit. From 2003, the Children’s Tax Credit was subsumed in CTC.

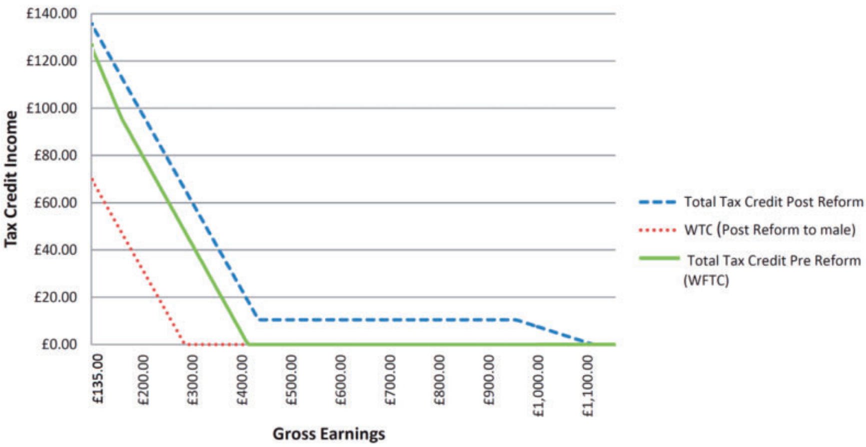

Fig. 1. Total tax credit income (weekly) for a single-earner couple with two children, husband working 30 hours per week
Source: Author's calculations based on Appendix Tables 2.1.2 and 2.1.3; x-axis truncated at 30 x national minimum wage = £135.

comparing the pre-reform 'WFTC' schedule to the 'Total Tax Credit Post-Reform' schedule. The amount of the post-reform tax credit income that is paid to the male partner can be seen from the WTC schedule. For example, we see that for a couple with gross earnings of £300 per week, WTC income is zero and the full total tax credit (of approximately £60) is paid via CTC. Note that the long-tail of the new tax credit schedule, which occurs as a small component of CTC (known as the family element), is not withdrawn until gross earnings reach a higher threshold. A full breakdown of how the maximum amounts available in WFTC were allocated to the new tax credits is available in the supplementary material (Tables 2.1.2 and 2.1.3).

Lastly, there is a smaller group of low-income families who do not satisfy the tax credit hours of work conditions. Prior to 2003, such families could instead claim equivalent payments for their children, which formed part of distinct means-tested benefits.[5] The couple would choose which partner received these payments (so the mother received payment only with the father's consent), but from 2003 they were subsumed in CTC. This implies that the group of tax-credit ineligible families also experienced an intra-household transfer of income from fathers to mothers as a result of the 2003 reform. This group of families constitute 11.7% of the treated group of households in our estimation sample.

3. Methodology

This section details the empirical approach of the paper. We are interested to know if paying state benefits to women in couples (instead of men) affects household spending patterns, and we use the 2003 UK tax-credit reform to give an exogenous change in the recipient of tax credits in couples. In particular, we are interested in estimating effects separately for individual spending items that are assignable to i) children, ii) adults and iii) collectively consumed goods.

5 These benefits are Income Support and Income-based Job Seeker's Allowance.

The key challenge in identifying the effect of the reform is separating its impact from other time-varying factors that may influence household spending such as price changes, income shocks, or changes in tastes. Our approach to addressing this problem is to make use of the commonly-used difference-in-differences (DID) estimator. We compare changes in expenditure of low income households that are more likely to have been affected by the reform (treated group) with changes in expenditure for middle-income households who are less likely to have been affected (control group). This approach rests on the fact that out-of-work benefits and in-work tax credits are targeted at low-income families and are reduced as household income increases. Taking the distribution of male take-home pay in a given year, the treated group is defined as households in the bottom quartile, and the control group as those in the inter-quartile range.[6]

Note that we use the treatment/control terminology in a somewhat extended sense as the exact level of earnings at which a families' tax credit entitlement reaches zero is family-specific (depending on family size, hours of work, childcare costs), meaning that some families assigned to the control group actually receive small amounts of tax credit and other families assigned to the treated group will in fact be ineligible for tax credits. Our results thus derive from differences in the intensity of treatment that each group receives.

As household expenditures follow the growth rate of the economy (i.e., growing proportionally) we specify a non-linear DID model for each expenditure item (discussed below) of interest:

$$y_i = exp(\alpha + \beta_1 d_i + \sum_{m=feb}^{m=dec} \beta_{2m} m_m + \sum_{t=2002}^{t=2005} \beta_{3t} t_t + \beta_4 d_i.I(t > 2002))\, \eta_i \qquad (1)$$

where y_i measures weekly household expenditure, d_i is a binary treatment group indicator ($d_i = 1$ for low income families), m_m is a set of binary month of interview indicators, t_t is a set of binary year of interview indicators, $I(t > 2002)$ is a binary variable indicating the post-reform period, and η_i is a mean one-error term. To interpret the coefficients, β_1 gives a constant mean difference between the treatment and control groups, the β_2's capture seasonal differences in expenditure, and the β_3's capture changes in expenditure due to sources other than the reform, including changing prices, tastes, etc. The critical DID identifying assumption is that, in the absence of the reform, the treatment and control groups follow the same spending trends in proportional terms. That is, we assume that both the treatment and control groups respond equally (in proportional terms) to price, income, and taste shocks. Under this assumption, $e^{\beta_4} - 1 \approx \beta_4$ is the treatment effect of interest and gives the constant percentage increase in household expenditure due to the reform.

To account for the possibility that there may be differences in group-specific treatment trends, we include a rich set of controls in the model, including controls for household composition and household income. The main results turn out to be relatively unaffected by the

6 Our methodology requires that male earnings be exogenous to the reform. This matter is discussed in Section 6.3. The treatment control/threshold (bottom quartile) corresponds to the income level at which tax credit entitlement reaches zero for a family with characteristics at the sample means (number of children, childcare costs, and assuming eligibility for the 30-hour component of tax credits). Households in the top quartile of the distribution are likely to differ from the treated group considerably and are dropped from the analysis. We experiment with the definitions of the groups in robustness checks (see Appendix 1). The groups were revealed to be stable in their composition under the main definition (full results are included in Appendix 2).

inclusion of the control variables. The full list of controls is discussed in Section 4. We further implemented a number of robustness checks, the results and details of which are summarized in Appendix 1. The robustness checks test the sensitivity of our results to the following: the common trends assumption, the possibility that the groups differ in their response to price and income shocks from a common proportional change, and the definition of the treatment and control groups.

Further potential threats to DID designs are as follows: i) implementation/ transitional problems, ii) other reforms taking place in the estimation period, and iii) anticipation effects. We argue that these are not important for the reform in this paper, and full details are documented in Appendices 3 and 4.

3.1 Estimation and inference

Standard practice is to log linearize eq. (1) and estimate it by ordinary least squares (OLS). However, the logarithmic transformation of eq. (1) raises two issues in the estimation of the log linear model by OLS (Santos-Silva and Tenreyro, 2006). First, observations for which expenditure is zero are dropped from the estimation sample, and this sample selection represents a source of bias.[7] Second, under heteroskedasticity the parameters of log-linearized models estimated by OLS lead to biased estimates of the true elasticities. To address the above, Santos-Silva and Tenreyro (2006) argue for using the Poisson Pseudo Maximum Likelihood estimator (PPML) to directly estimate the model in non-linear form. Direct estimation by PPML means that zero observations can be kept, and requires only mean independence of the error term η_i (and not statistical independence). We therefore estimate eq. (1) directly by PPML. The reform could lead to changes in both the levels and proportions of expenditure devoted to various expenditure categories. For this reason, alongside the preferred PPML estimates, results are also presented from a linear specification for expenditure shares (estimated by Ordinary Least Squares).

In terms of inference, regular standard errors may overstate the precision of estimates of a treatment effect in DID designs. Procedures to correct for this have been the subject of fierce debate and the literature is still unsettled on how to proceed in the one-treatment-one-control group case (Wooldridge, 2003; Donald and Lang, 2007). Bertrand *et al.* (2004) show that the main source of bias arises from serial correlation. The authors present evidence that the bias is largely eliminated when focusing on short time spans.

Acknowledging the above, we proceed by restricting the sample to five periods to address serial correlation concerns and report heteroskedasticity robust standard errors. To give further credibility to our results, we estimated placebo effects for all spending items. None of the estimated placebo interventions were statistically significant, which we interpret as evidence that within-group shocks are not a problem for inference (for details, see robustness check one, Appendix 1). This approach is in the spirit of Abadie *et al.* (2010), where the authors are able to calculate the exact distribution of treatment effects from random placebo interventions.

3.2 Choosing spending items

Our research strategy involves testing for reform effects in multiple spending items consumed by: i) children, ii) adults, or iii) collectively. For children's goods, we consider goods that have been used as measures of parental child investments in the literature. We expect

7 For example, in our data 62% of households report zero weekly expenditure on male clothing.

positive reform effects for the children's goods if they are more strongly female preferred. We consider children's clothing, fresh fruit and vegetables, childcare, books, newspapers and magazines, musical instruments, toys, and games. For the private adult goods, we follow LPW in choosing the following: clothing expenditures; cosmetics that we conjecture are more extensively consumed by women in the household; takeaway meals that may represent a substitute for home production; gambling, where women are less likely to make risky choices than men; and maintenance payments, which, if children typically reside with the mother following divorce, are paid in respect of the male partner's children from a former marriage, and thus reflect the male partner's preference.

In line with the above discussion, we also contribute two new public goods that are collectively consumed: spending on holidays and home improvements. Here the expected signs are less clear. However, we conjecture that the partners of a couple will have different incentives about which of these goods should be provided. In particular, if the male partner attaches some positive probability to the possibility of relationship breakdown, he would prefer current resources to be spent on collectively-consumed goods where the benefit is immediate (i.e., a holiday) rather than on home improvements from which a large part of the benefit will be realized when he may no longer be a member of the household.

One concern is that when testing for reform effects in multiple spending categories, we would expect a few of the effects to be statistically significant by chance. Romano *et al*. (2008) review statistical approaches for dealing with multiple hypothesis testing, and comment that methods that control the probability of making at least one type one error (Family Wise Error Rate (FWE)) may be 'playing it too safe'. That is, we may prefer to use standard p-values and to accept the increased risk of making a type-one error for the benefit of not missing important treatment effects where they occur. This is a popular approach.

To address concerns over multiple comparisons, we argue that we have a strong theoretical basis for expecting reform effects of a particular sign, as discussed above. Finally, whilst we present unadjusted p-values in the main tables, Appendix 1 summarizes results from controlling the False Discovery Rate (FDR) by applying the Benjamini *et al*. (2006) procedure across all expenditures for which effects are estimated. Given that our sample sizes are relatively small and we give an upfront justification for the selected spending items, we control the FDR at 20% (details of implementation are included in Appendix 5).

4. Data

This paper uses expenditure data on couples by pooling the first five years of the Expenditure and Food Survey (EFS). The estimation sample consists of single-couple households (married or cohabiting) with at least one child aged 0–15, and responding to the EFS in one of the first five years of the survey (2001–2005).[8] The EFS operates on the basis of a financial year (April–March). The sample is, therefore, made up of two pre-reform and three post-reform years of data. It is further restricted to households where both partners are aged 24–59, not sick or injured, not self-employed, and not in full-time education. EFS interviews take place over a year's time and all income and expenditure figures are expressed in December 2005 terms by applying the all-items retail price index, available from the Office for National Statistics.

8 Households with children aged 16–18 were subject to Educational Maintenance Allowance reforms in 2004 and are excluded from the analysis.

Imposing the complete set of exclusions and the treatment/control group definitions in Section 3 leaves a baseline sample of 3,757 married or cohabiting couples with children: 1,257 in the treated group, and 2,500 in the control group.

Table 1 presents characteristics of interest for each group. Statistically significant differences are observed. Households in the low-income treatment group tend to be younger (0.87 years for men, 1.38 for women), lower-educated (0.57 years for men, an insignificant difference for women), have larger families (0.12 extra children), less likely to be married (11 percentage points), and considerably more likely to be in social housing (20 percentage points). The differences in region of residence (not presented) are generally small and statistically insignificant. Insofar as expenditures are likely to be affected by family demographics, we include a full set of controls in the main regressions. We include controls to capture differences in expenditure by the following: the number of children by age and sex (continuous variables for the number of children of each sex aged 0–1, 2–4, 5–15); parental age (linear terms plus square and cubic terms); parental education (dummies for male and female education, i.e., low, medium, or high); a continuous household income variable, an indicator for living in social housing, and 12 regions of residence dummies.

Table 1. Summary statistics by treatment and control group

Controls	*Low-income treated*[‡]	*High-income control*[§]	*Difference*
Age:			
Male partner	37.54	38.41	**−0.87*****
Female partner	34.90	36.28	**−1.38*****
Age left education:			
Male partner	16.82	17.40	**−0.57*****
Female partner	17.33	17.54	**−0.20**
Number of children:			
Total	1.93	1.80	**0.12*****
Age 0–4	0.62	0.60	**0.02**
Age 5–15	1.31	1.21	**0.10****
Housing and marriage:			
Social housing	0.27	0.07	**0.20*****
Married	0.76	0.87	**−0.11*****
Labour market:			
Household income [\\ ¶]	435.27	613.22	**−177.95*****
Employed (Male)	0.79	0.99	**−0.21*****
Employed (Female)	0.64	0.80	**−0.17*****
Work hours (Male)	30.46	40.12	**−9.66*****
Work hours (Female)	18.17	21.10	**−2.93*****
Number of households:	1,252	2,505	

Notes:

[†]Standard errors(robust) in parentheses. *p < .10, **p < .05, ***p < .01.

[‡]Treated group formed of households in the bottom quartile of the year-specific net male income distribution.

[§]Control group formed of households in the inter-quartile range of the year-specific net male income distribution.

[\\] £per week (December 2005 prices).

[¶]Wages + investment income + social security benefits - taxes.

The main reform effects of interest correspond to the spending items discussed above. Table 2 shows the pre-reform expenditure means for the spending items grouped according to private adult goods, children's goods, and collectively-consumed goods. For an examination of treatment and control trends for the spending items centered at the time of the reform, see Figs 2.1 and 2.2 in Appendix 2.

5. Comparison to 1970s sample

LPW exploit changes to the UK Child Benefit in the 1970s to demonstrate that redirecting transfer income from men to women in couples had an impact on expenditure benefiting children and women. Prior to 1977, child amounts in the benefit system were paid through a tax allowance (typically meaning higher take-home pay to the father). This allowance was phased out and replaced in 1979 with a non-taxable payment made directly to the mother. LPW find that expenditure on children's as well as women's clothing rose relative to spending on men's clothing. WB confirms the clothing findings, and further finds that the fraction of income devoted to toys, pocket money, and restaurant and takeaway meals all increase, whilst a 'men's tobacco' category (consisting of cigars, pipe tobacco, and snuff products) sees a decrease.

Table 2. Pre-reform summary statistics by treatment and control group

	Low-income treated†				High-income control‡			
Pre-reform expenditure	*Level*		*Share*		*Level*		*Share*	
Private adult goods								
Women's clothes	9.37	(18.60)	0.0191	(0.034)	12.66	(23.11)	0.0193	(0.030)
Men's clothes	6.01	(15.49)	0.0123	(0.030)	8.48	(19.71)	0.0129	(0.029)
Cosmetics	7.25	(9.39)	0.0158	(0.018)	9.99	(10.89)	0.0165	(0.017)
Takeaway meals	5.70	(7.51)	0.0138	(0.019)	6.16	(7.57)	0.0108	(0.014)
Gambling	2.99	(6.12)	0.0069	(0.014)	2.87	(5.67)	0.0052	(0.011)
Maintenance payments	1.91	(11.07)	0.0042	(0.023)	2.08	(12.14)	0.0032	(0.018)
Child-related goods								
Children's clothing	8.70	(14.93)	0.0189	(0.029)	10.20	(14.38)	0.0174	(0.025)
Fresh fruit/vegetables	4.18	(3.80)	0.0106	(0.012)	5.54	(4.29)	0.0099	(0.008)
Childcare	4.23	(17.97)	0.0073	(0.029)	10.88	(39.68)	0.0146	(0.047)
Books	4.13	(4.91)	0.0098	(0.012)	5.16	(6.64)	0.0089	(0.010)
Toys	4.33	(10.73)	0.0091	(0.019)	6.45	(13.27)	0.0104	(0.022)
Musical instruments	0.12	(1.48)	0.0002	(0.003)	0.62	(6.56)	0.0009	(0.009)
Collectively consumed household goods								
Holiday	3.34	(11.24)	0.0069	(0.022)	5.31	(14.78)	0.0084	(0.023)
Home improvements	1.31	(9.31)	0.0026	(0.018)	7.89	(96.79)	0.0059	(0.044)
Number of households:		527				1052		

Notes:
*Standard deviations in parentheses.
**Levels expressed in £per week (December 2005 prices).
†Treated group formed of households in the bottom quartile of the year-specific net male income distribution.
‡Control group formed of households in the inter-quartile range of the year-specific net male income distribution.

We calculate that the average CTC claim for a two-child family in our sample represents approximately 8% of total weekly household expenditure in 2003. It is interesting to point out that this figure lines up closely with the WB and LPW 1970s samples, where the intra-household transfer in their papers corresponds to 6% of the average total household expenditures (based on calculations from Ward-Batts, 2008).

One difference with the 1970s reforms is that Child Benefit payments are universal, whereas the 2003 reform targeted low-income families only. LPW and WB both estimate the overall effect of the 1970s reforms, but their results do not tell us if the magnitude of the treatment effect differs across subgroups. If low-income households showed the strongest response to the reforms, then the results from the earlier studies would represent a lower bound for the effect of the reforms on low-income households' expenditures. As such, the difference between the findings of the earlier studies and the results in this paper could be interpreted as a lower bound for the change in the effects on low-income households.

It is informative to compare the characteristics of our treated group of couples to those reported by WB for the 1970s sample. Table 3 confirms that there have been considerable changes in the social and economic setting in the UK since the 1970s. Couples in the current paper are substantially more likely to have a female in employment (64% vs. 48%) but are similar in terms of ages. Conditional on having at most three children, the 1970s sample has a larger share of two- and three-child families. In the current paper, we observe higher

Table 3. Comparison of sample to WB's 1970s sample

	1970s sample	2001–2005 sample (treated group)
Female employed	48.2	64
Male age	36.5	37.54
Female age	33.9	34.9
Number of Children:		
1 child	33.4	37.38
2 child	48.4	41.37
3 child	18.2	14.86
3+ children	–	6.39
N	15,753	1252
Expenditure ratios†		
Children/men's clothing:		
One child	0.97	1.23
Two child	1.63	2.46
Three child	2.2	2.74
Women's/men's clothing:		
One child	1.7	2.35
Two child	1.6	3.82
Three child	1.77	1.78
Gender wage ratio‡	69	76

Notes:
†Means from LPW table 3.
‡Figures from Myck and Paull (2004). Ratio of average hourly wages for full time workers (female/male).

ratios of child to male clothing expenditure across all three family sizes reported. For example, for a two-child family the ratio increased from 1.63 in the 1970s to 2.46 at the start of the 21st century. Similarly, the ratio of women's to male clothing has increased for one- and two-child families. Also, the table reports differences in the gender average hourly wage ratio (women's wage/men's wage). The ratio has increased from 69% in 1979 to 76% in 2000.

Taken together, these simple descriptive statistics show that there have been important changes within couples across the two compared periods, and the starting position of women and children in 2003 appears to be stronger than in the 1970s.

6. Results

6.1 Private adult and collectively consumed household goods

Panel A of Table 4 presents the estimated reform effects for the six private adult goods, whereas panel B shows estimates for the two collectively consumed household goods. In order to explore the effects of receiving a greater intensity of treatment, columns 5–8 remove the top-earning one-fifth of the treated group from the estimation sample. While this definition of the treated group focuses better on the most affected households, the associated cost is that the treated and control groups now differ more in terms of their observable and unobservable characteristics. Results in line with the main findings would, however, be reassuring.[9]

Starting with the private adult goods, reassuringly, we see coefficient estimates that are stable in sign across the specifications and estimation methods. Evidence of important reform effects is found amongst the parent goods. For gambling payments, the preferred PPML estimates in column 1 show a statistically significant reduction in gambling expenditure. The effect becomes more precisely estimated when the household-level controls are added to the model, and implies a reduction in spending due to the reform of -34.2%, calculated as $exp(\widehat{\beta}_4) - 1$. A statistically significant effect of a similar magnitude is found for the model with controls in the smaller high-treatment intensity sample, although the effects for expenditure shares, while of the same sign, are never statistically significant.

Maintenance payments typically represent a transfer from men to their former partners and children, and as such the new female partner is likely to prefer these to be lower, whereas the male partner may wish to continue to support his children from his former relationship (and perhaps his former partner). We find results in line with this interpretation where we see that the treated group reduces spending in this area, relative to the control, after the introduction of the new tax credits. Coefficients are consistently negative across the columns and become larger in absolute terms and statistically significant in columns 5–8 when focusing on the most intensely treated households. The findings are robust to the inclusion of the control variables. The PPML results with controls in column 2 imply a reform effect of -40.6%. To put the magnitude of this effect into context, Ermisch and Pronzato (2008) find in their sample of re-partnered British men that a 10 percentage point increase in the female income share reduces the share of household income devoted to maintenance payments by 0.0037.

9 A similar robustness check is performed that removes the lowest earning households from the control group (for details, see robustness check 2, Appendix 1.)

Table 4. Estimates of the reform effect on private adult and collectively consumed household goods

	Full sample				High-treatment intensity sample			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable:	PPML	+ controls	Shares	+ controls	PPML	+ controls	Shares	+ controls
A. Private Adult Goods								
Women's clothes	0.104	0.136	0.001	0.002	0.065	0.094	0.001	0.001
	(0.130)	(0.128)	(0.002)	(0.002)	(0.137)	(0.135)	(0.002)	(0.002)
Men's clothes	0.042	0.046	0.001	0.001	0.058	0.059	0.001	0.001
	(0.173)	(0.171)	(0.002)	(0.002)	(0.194)	(0.190)	(0.002)	(0.002)
Cosmetics and related	−0.065	−0.041	−0.001	−0.001	−0.055	−0.031	−0.001	−0.001
	(0.086)	(0.085)	(0.001)	(0.001)	(0.097)	(0.095)	(0.001)	(0.001)
Takeaway meals	0.118	0.122	0.002	0.002	0.140	0.137	0.002	**0.002***
	(0.088)	(0.087)	(0.001)	(0.001)	(0.094)	(0.093)	(0.001)	(0.001)
Gambling	**−0.515***	**−0.418****	−0.001	−0.001	−0.456	**−0.380***	−0.000	−0.000
	(0.285)	(0.191)	(0.001)	(0.001)	(0.298)	(0.201)	(0.001)	(0.001)
Maintenance or separation payment	−0.576	−0.521	−0.002	−0.002	**−0.937***	**−0.853***	**−0.003***	**−0.003***
	(0.428)	(0.417)	(0.001)	(0.001)	(0.499)	(0.478)	(0.002)	(0.001)
B. Collectively consumed household goods								
Holiday	−0.301	−0.318	−0.003	−0.002	**−0.568****	**−0.555****	**−0.004****	**−0.003****
	(0.234)	(0.224)	(0.002)	(0.002)	(0.257)	(0.243)	(0.002)	(0.002)
Home improvements	**2.604*****	**2.473*****	**0.006****	**0.005****	**2.601*****	**2.422*****	**0.006****	**0.006***
	(0.707)	(0.637)	(0.003)	(0.003)	(0.773)	(0.658)	(0.003)	(0.003)
Observations	3,757	3,757	3,757	3,757	3,509	3,509	3,509	3,509

Notes: Standard errors (robust) in parentheses. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. Models include a full set of year and month of interview dummies, a disposable household income variable, a treatment status indicator and a post-reform indicator x treatment status interaction (coefficient presented). Controls are as follow: a full-set of region dummies; variables for the number of male and female children residing in the household in age categories: 0–1, 2–4, and 5–15; an indicator for residing in social housing; controls for parental characteristics: age, age squared, age cubed, and age left full-time education indicators (greater than age 16 and less than 21 years old; greater than 21 years old).

The central result presented in the LPW study of the 1970s child benefit reforms was statistically significant increases in the ratio of women's to men's clothing. We are therefore particularly interested to see how adult clothing expenditure changed following the reform at the start of the 21st century. For women's clothing, we see small positive but highly insignificant reform effects for both the preferred PPML results and models for the share of expenditure. Furthermore, when focusing on the most affected households, the PPML estimates become smaller in magnitude and remain highly insignificant. The effects for male clothing are also of the wrong sign (positive) and highly insignificant. To contrast this with WB, whose estimated effects imply an increase in female clothing expenditure of approximately +22% and a –43% fall in male clothing expenditure, our respective estimates are +14.5%, and +4.7%.

One explanation for the contrast with the 1970s reforms is the dramatic fall in the price of clothing seen in recent decades, potentially making decisions regarding clothing expenditure less contentious.[10] Also, changed shopping patterns may have contributed to clothing expenditure becoming less contentious. Whereas in the 1970s clothing and footwear would be purchased in specialist shops requiring a journey to a town centre and the participation of both parents, by the early 2000s low-income households would be accustomed to buying clothing and footwear at the supermarket as part of a regular shopping trip. If women were already making most of these expenditure decisions prior to 2003, then the reforms might be expected to have little impact on expenditure patterns on these items.[11]

Interestingly, the estimated signs for both cosmetics and takeaway meals are in line with the earlier child benefit results of Ward-Batts. The estimates of the tax credit reform in this paper are, however, small and statistically indistinguishable from zero, excluding the takeaway meals results for expenditure shares for the high-treatment intensity sample. One explanation for the smaller effects for takeaway meals in this paper could be the rise of cheap, quick-to-prepare meals and pre-prepared vegetables. This has reduced the time inputs required in the production of home cooked meals and has created a cheap alternative to takeaway meals.

One concern is that while more resources are predicted to be devoted to the most strongly preferred female goods (such as children's goods), offsetting spending reductions in male-preferred public goods may still impose a child development cost, if they are goods which the whole family consumes. We therefore extend the previous literature to consider two such goods, spending on the home and spending on holidays, in panel B of Table 4.

For holiday spending, negative effects are found in the PPML models with and without controls. While the effects are statistically indistinguishable from zero in the main sample, they grow in absolute magnitude when focusing on the most highly-treated households in columns 5–8, and become statistically significant. The PPML results in column 2 imply a decrease in holiday spending due to the reform of -27.2%. The inclusion of controls makes little difference to the estimated coefficients and tends to improve the precision somewhat. The pattern for the spending shares reinforces the finding with negative and statistically significant effects in the restricted sample, both with and without controls.

10 For example, whilst clothing and footwear expenditure accounted for approximately 12% of household expenditure in LPW, the same figure for this paper is 5%.

11 An alternative explanation could be a lack of statistical power, although we are able to detect reform effects in a number of other spending areas that follow.

For the final collectively consumed good, particularly strong effects are found. The PPML estimates in column 2 imply an increase in spending on home improvements of 108.6%. The estimate is robust to the inclusion of the controls, falling a little in magnitude, but remaining statistically significant at the 1% level. The high treatment intensity sample gives further support to this finding, with the preferred estimate implying a statistically significant effect of a similar magnitude. Finally, the results for the expenditure shares also line up, and we see positive statistically significant reform effects in both samples, which are again robust to the inclusion of controls to the model.

A way to make sense of the findings for holidays and home improvement expenditure follows from the discussion in Section 3.2, where couples face a positive probability of divorce. The male partner puts less weight on spending from which a large part of the benefits accrue in a period in which he may not be a member of the household (home improvements) and puts greater weight on spending where benefits are realized immediately (holiday spending).

In summary, the evidence presented in Table 4 indicates that redirecting tax-credit income from fathers to mothers leads to important changes to the quantity of resources that low-income households devoted to specific items of expenditure. In terms of the private adult goods, we see evidence of reductions in expenditure on gambling and maintenance payments, but interestingly and in contrast to earlier reforms, no effect for adult clothing expenditures. The new evidence presented here suggests a trade-off in expenditure between goods that are consumed by the whole family—increases are observed for spending on home improvements but reductions are seen in holiday spending. The latter result could be rationalized if the incentives to invest in the different collectively-consumed household public goods vary by sex.

6.2 Child spending

We are particularly interested to know how directing benefit income to women may affect spending on children's goods. Table 5 repeats the empirical exercise of the previous section for the child spending items. The table indicates that following the 2003 tax credit reform, low income households increased their expenditures relative to the control group in some areas of child spending. Both the PPML results for the full sample and the models for expenditure shares imply an increase in weekly expenditure on toys and games that is statistically significant. Further strength is given to the finding with the inclusion of controls to the models where the effects remain statistically significant. For the smaller high treatment intensity sample, the estimates are of a similar magnitude excluding the PPML with controls, where the effects are slightly larger but less precisely estimated. As with some of the earlier results for the private adult goods, the finding for toys and games mirrors that of WB for the 1970s child benefit reforms, where positive spending increases on toys were found of a similar magnitude (approximately 29% in both WB and our estimate in column 2).

Cunha and Heckman (2008) consider ownership of musical instruments as a measure of parental child investments. We estimated separate reform effects for musical instrument spending and find a result that lines up with the toys and games finding. That is, we find evidence of statistically significant spending increases on musical instruments following the introduction of the new tax credits in 2003.

To comment on the remaining items in the table, for fruit and vegetables, the preferred PPML estimates are positive, but never statistically significant. For the remaining items, the

Table 5. Estimates of the reform effect on child goods

Dependent variable:	Full sample (1) PPML	(2) + controls	(3) Shares	(4) + controls	High treatment intensity sample (5) PPML	(6) + controls	(7) Shares	(8) + controls
Child Goods								
Children's clothing	−0.066	−0.052	0.001	0.001	−0.048	−0.023	0.001	0.001
	(0.109)	(0.107)	(0.002)	(0.002)	(0.119)	(0.116)	(0.002)	(0.002)
Fresh fruit/ vegetables	0.079	0.062	0.000	0.000	0.053	0.036	0.000	0.000
	(0.060)	(0.055)	(0.001)	(0.001)	(0.066)	(0.060)	(0.001)	(0.001)
Childcare	0.166	0.068	0.001	0.000	0.117	−0.002	0.000	−0.001
	(0.305)	(0.297)	(0.003)	(0.003)	(0.355)	(0.349)	(0.003)	(0.003)
Books, news-papers, magazines	−0.036	−0.024	−0.000	−0.000	−0.040	−0.024	−0.000	−0.000
	(0.088)	(0.085)	(0.001)	(0.001)	(0.097)	(0.093)	(0.001)	(0.001)
Toys and games etc.	**0.269***	**0.255***	**0.003***	**0.003***	**0.309***	0.278	**0.003***	**0.003***
	(0.157)	(0.155)	(0.001)	(0.001)	(0.173)	(0.171)	(0.002)	(0.002)
Musical instruments	**1.800****	**1.795****	**0.001***	**0.001***	**1.780****	**1.888****	**0.001***	**0.001****
	(0.861)	(0.777)	(0.000)	(0.000)	(0.874)	(0.778)	(0.000)	(0.000)
Observations	3757	3757	3757	3757	3509	3509	3509	3509

Notes: See Table 4 notes.
*p < .10, **p < .05, ***p < .01. Standard errors (robust) in parentheses.

estimates are highly insignificant and not suggestive of an effect on household spending patterns. This includes the case of children's clothing, and matches with the result of no reform effect on adult clothing from the previous section. WB found an effect of approximately 24% for children's clothing expenditure, whereas our estimate is a statistically insignificant -5%. As documented in Section 5, the ratio of child (and women) to male clothing spending in the sample period of this paper is much more favorable to children (and women) than in the period of the 1970s reforms, and taken together with the estimated effect here, it suggests that clothing spending decisions are less contentious in the current period.

In summary, this section has presented evidence that introducing CTC leads low-income households to allocate greater resources to important areas of child spending but not for children's clothing. Despite this difference with the LPW and WB studies, the picture emerging is similar, with spending increases found for toys and games. We also find increases in spending on musical instruments.

6.3 Female labour supply

We have so far argued that redirecting benefit income from males to females is the key feature of the 2003 reform. But if there were labour supply responses to the reform, then this could provide an alternative explanation for the observed changes in spending patterns. Previous studies suggest that male labour supply is insensitive to changes in financial incentives (see Meghir and Phillips 2010) but female labour supply can be responsive. No published work has estimated the labour market impact of the new tax credits, but we point to three pieces of evidence against a female labour supply interpretation.

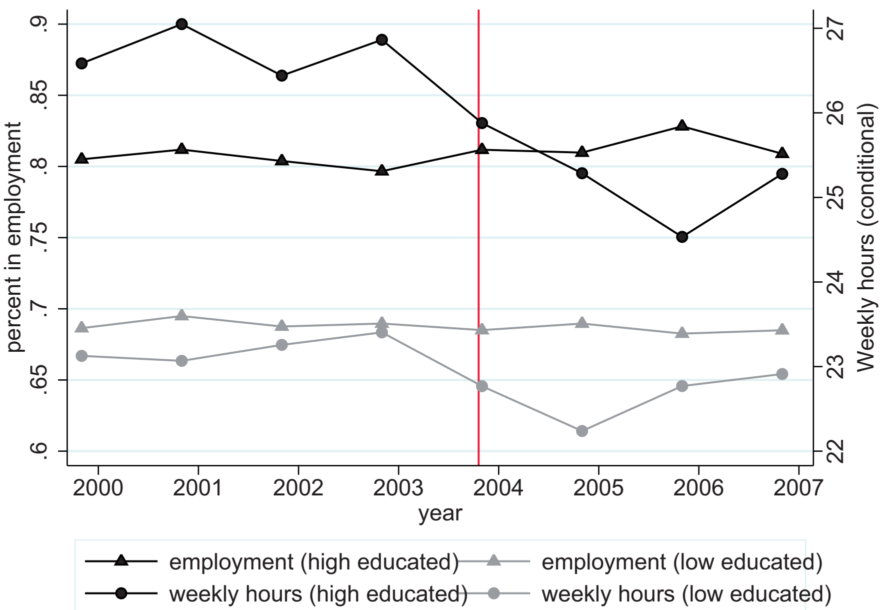

Fig. 2 Labour market trends for females in couples with children
Source: Labour Force Survey 1999–2006 (Oct–Dec). Sample: women in couples with dependent children. Highly-educated: highest qualification is at least a degree or equivalent. Low educated: highest qualification is less than a degree or equivalent.

First, the reform did not target labour supply, and the incentive changes involved are very small (see the changes in total tax credit income in Fig. 1). Brewer (2003) looks at the impact of the reform on work incentives and comments that '. . . . . . low-earning families with children generally receive higher incomes both in and out of work. This improves the financial gain to working compared with not working for some and reduces it for others. Most of these changes, though, are relatively small in magnitude.' Indeed, the fact that there was no expected labour market response may explain the absence of any empirical work estimating the labour market impact of the reform.

Second, using the UK Labour Force Survey (LFS), Fig. 2 presents female labour market trends centered around 2003. The figure shows trends in employment and hours of work by education level for females in couples with children. If the reform had a female labour market impact then we would expect this to show up for the lower-educated group (who are more likely than the higher-educated to be receiving state benefits and so more likely to be affected by the reform) following 2003. The employment rate for the low-educated women is stable at around 69% across the full period (2000–2007) and there are no obvious discontinuities in the series at 2003. We do observe a small fall in weekly work hours for the low-educated women from a mean of 23.2 hours in the pre-reform years to 22.7 hours in the post-reform period. However, this small fall in hours of work is matched for the group of highly-educated women (who are much less likely to receive tax credits), suggesting that this change is not driven by the reform but rather by macro economic factors affecting both groups equally.

Third, our methodology allows us to directly test for responses in female labour supply and hours of work, and the estimated effects are small and highly insignificant. Our estimated employment effect is 0.019 with a standard error of 0.030; for weekly hours worked the estimated effect is -0.136 with a standard error of 0.112, which we interpret as evidence against a female labour supply response.

7. Conclusions

This paper has empirically examined whether paying welfare benefits to women instead of men can influence household spending; it does this by exploiting a UK reform in 2003 that caused an exogenous transfer of tax credit income in couples from men to women. Our primary finding is that the reform caused low-income households to reallocate spending.[12]

LPW, reprised in WB, consider a UK child benefit reform in the late 1970s that caused a wallet-to-purse transfer of a similar magnitude to the reform examined in this paper, which took place at the start of the 21st century. LPW find that the ratio of women's to men's, and children's to men's clothing expenditures are both found to increase following the child benefit changes. We find clothing effects that are much smaller in magnitude and statistically insignificant. Indeed, we note that the ratio of women (and child) to male clothing spending is higher in the current paper, potentially meaning that household decisions regarding clothing expenditure have become less contentious. One implication of this finding is that research that has traditionally relied on children's and female clothing expenditures as indicators of female spending power will become more difficult over time.

12 The UK Government is planning a further round of welfare simplification with the aim of a single 'Universal Credit' replacing six existing benefits, including WTC/CTC, by 2017. In couples, the Universal Credit will be paid to a 'nominated person', not necessarily the mother.

Our second key finding, which is consistent with the 1970s results of WB, is that directing benefits to women increases spending on children (toys and games, musical instruments). This evidence points to a model of household decision making that must deliver the feature of greater spending on children's goods when income is placed in the hands of women, and this must be robust to the observed changes over time in female labour force participation.

Our third result is of an apparent trade-off in spending on goods that are collectively consumed. We find that the exogenous transfer of tax credit income from men to women leads to households increasing spending on home improvements, but decreasing spending on holidays. Insofar as these are two goods that are collectively consumed, it makes the overall impact of targeting benefits to women on child consumption less clear.

An interpretation of the housing and holiday spending findings rests on the fact that the gains from marriage may be captured by one person if the marriage breaks down (Stevenson 2007). The male partner, knowing that he may not be a member of the household in future periods, will prefer expenditures to be on items where the benefits are immediate (e.g., holidays) rather than on items from which the benefits may last for several years, that is, into periods in which he might no longer be a member of the household (e.g., home improvements). This interpretation suggests that research that looks for effects of intra-household transfers on household spending could use the collectively-consumed goods as an alternative to the adult and children's clothing expenditures, that as documented above, are less likely to be effected by intra-household transfers over time.

Supplementary material

Supplementary material (Appendices 2–6) is available online at the OUP website.

Funding

This work was supported by the Economic and Social Research Council.

Acknowledgments

I would like to thank Marco Francesconi, Mike Brewer, Joao Santos Silva, Susan Harkness, Emanuele Ciani, Iva Tasseva, and seminar participants at the Royal Holloway University of London and the European Society for Population Economics annual conference 2011 in Hangzhou, China for their useful comments. I am also very grateful to two anonymous referees whose suggestions led to substantial improvements to the paper. Data from the Expenditure and Food Survey has been made available by the Office for National Statistics (ONS) through the UK Data Archive, and has been used with permission. Neither the ONS nor the Data Archive bear any responsibility for the analysis or interpretation of the data reported here.

References

Abadie, A., Diamond A., and Hainmueller, J. (2010) Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, 105, 493–505.

Attanasio, O. and Lechene V. (2010) Conditional cash transfers, women and the demand for food. IFS Working Papers W10/17, Institute for Fiscal Studies, London.

Benjamini, Y., Krieger, A.M., and Yekutieli, D. (2006) Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93, 491–507.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004) How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119, 249–75.

Bobonis, G.J. (2009) Is the allocation of resources within the household efficient? New evidence from a randomized experiment. *Journal of Political Economy*, 117, 453–503.

Braido, L.H.B., Olinto, P., and Perrone, H. (2012) Gender bias in intrahousehold allocation: Evidence from an unintentional experiment. *The Review of Economics and Statistics*, 94, 552–65.

Brewer, M. (2003) The new tax credits. Technical report, IFS Briefing Notes, Institute for Fiscal Studies, London.

Cunha, F. and Heckman, J.J. (2008) Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, 43, 738–82.

Donald, S.G. and Lang, K. (2007) Inference with difference-in-differences and other panel data. *The Review of Economics and Statistics*, 89, 221–33.

Duflo, E. (2000) Child health and household resources in South Africa: evidence from the old age pension program. *American Economic Review*, 90, 393–98.

Duflo, E. and Udry, C. (2004) Intrahousehold resource allocation in Cote d'Ivoire: social norms, separate accounts and consumption choices. NBER Working Paper, Cambridge, MA.

Ermisch, J. and Pronzato, C. (2008) Intra-household allocation of resources: Inferences from non-resident fathers' child support payments. *Economic Journal*, 118, 347–62.

Gregg, P., Waldfogel, J., and Washbrook, E. (2006) Family expenditures post-welfare reform in the UK: are low-income families starting to catch up? *Labour Economics*, 13, 721–46.

Hotchkiss, J.L. (2005) Do husbands and wives pool their resources? Further evidence. *Journal of Human Resources*, 40, 519–31.

Inland Revenue. (2003) Child and Working Tax Credit quarterly statistics October 2003. Technical report, Inland Revenue, London.

Lundberg, S.J., Pollak, R.A., and Wales, T.J. (1997) Do husbands and wives pool their resources? Evidence from the United Kingdom child benefit. *Journal of Human Resources*, 32, 3463–80.

Meghir, C., and Phillips, D. (2010) Labour supply and taxes. In J. Mirrlees, S. Adam, T. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. Myles, and J. Poterba (eds), *Dimensions of Tax Design: The Mirrlees Review*, Oxford University Press for Institute for Fiscal Studies, Oxford.

Myck, M., and Paull, D. (2004) The role of employment experience in explaining the gender wage gap. IFS Working Papers WP04/16, Institute for Fiscal Studies, London.

Romano, J.P., Shaikh, A.M., and Wolf, M. (2008) Formalized data snooping based on generalized error rates. *Econometric Theory*, 24, 404–47.

Santos-Silva, J.M.C., and Tenreyro, S. (2006) The log of gravity. *Review of Economics and Statistics*, 88, 641–58.

Stevenson, B. (2007). The impact of divorce laws on marriage specific capital. *Journal of Labor Economics*, 25, 75–94.

Ward-Batts, J. (2008). Out of the wallet and into the purse: using micro data to test income pooling. *Journal of Human Resources*, 43, 325–51.

Wooldridge, J.M. (2003). Cluster-sample methods in applied econometrics. *American Economic Review*, 93, 133–38.

Appendix 1: robustness checks

This appendix contains details of seven robustness checks performed, and Table A.1 summarizes the results of these checks. Full tables of results from the robustness checks are available in the supplementary material (Appendix 6).

Table A1. Summary of statistically significant effects in robustness checks

Robustness Check:	1	2	3	4	5	6	7
Parent Goods							
Women's Clothes	×						
Men's Clothes	×						
Cosmetics and Related	×						
Takeaway Meals	×	✓			✓	✓	
Gambling	×	✓					✓
Maintenance Payments	×	✓					✓
General Household Goods							
Holiday	×	✓	✓	✓	✓	✓	✓
Home Improvements	×	✓	✓	✓	✓	✓	✓
Child Goods							
Children's Clothing	×						
Fresh Fruit/Vegetables	×						
Childcare	×						
Books, Newspapers, Magazines	×						
Toys and Games	×		✓	✓		✓	
Musical Instruments	×	✓	✓	✓	✓	✓	✓

Notes: ✓indicates a statistically significant reform effect (10% level); × indicates a statistically insignificant reform effect in the placebo reform. Models are the preferred PPML estimates with controls and from the high treatment intensity sample.

Robustness check 1—placebo reform: A concern that the main expenditure findings are driven by differences in unobservables across the male net earnings groups or are due to chance is further addressed by performing a placebo reform. The treated group consists of families in the second quartile of the male net earnings distribution who are unaffected by the reform; families in the top half of the male net earnings distribution form the control group.[13] To comment on the main findings, all of the estimated coefficients (PPML with controls) for the adult, child, and collectively consumed household goods are statistically insignificant. This is strong evidence against the hypothesis that the main results are driven by differences in trends across the control and treatment groups or are due to chance.

Robustness check 2—less treated control group: The reform effect is identified from differences in the intensity of treatment. To better focus the control group on households unaffected by the reform, this robustness check removes the lowest income households from the control group. The control groups is defined as households with male earnings between percentiles 30 and 75. The associated cost is that, under this definition, the treatment and control groups differ more in terms of their characteristics.

Robustness check 3—male partner in at least eligible employment: A sample restriction is imposed that attempts to remove households from the treated group, which does not experience a wallet-to-purse transfer. The sample is restricted to households in which the male partner is in tax credit eligible employment—that is, reports working at least 16 hours

13 Typically in a DID setting, it would be informative to perform placebo reforms outside of the sample period for the main construction of treatment and control groups. However, changes to the EFS survey and numerous other reforms taking place outside the main estimation period (e.g., the introduction of WFTC in 1999), mean that such tests are uninformative.

per week. In this way, households where the female has sole entitlement to receive tax credit payments are removed from the sample.

Robustness check 4—low-income control group: One may worry that the control and treatment groups do not experience the same proportional response to price and income shocks as assumed by the exponential specification. This robustness check restricts the control group to couples that are more similar in their income levels to the treated group. Specifically, the control group is restricted to include couples only in the second quartile of the male net labour income distribution. The cost of this restriction is a loss of sample size and precision.

Robustness check 5—excluding observations seven months prior to the reform: Applicants to WFTC after August 2002 received payment as a direct benefit payment, until the implementation of the reform in April 2003. To check the sensitivity of the results to this transition, observations in the seven months prior to the reform are dropped from the estimation sample.

Robustness check 6—differential linear trends: The main estimation sample includes two pre-reform and three post-reform years of data, raising the possibility of controlling for differential treatment/control trends, albeit in a restrictive manner. If the linear specification is the appropriate one, then including these terms should not affect the estimated reform effects.

Robustness check 7—multiple hypothesis testing adjustment: This test implements the Benjamini *et al*. (2006) procedure to control the false discovery rate at 20 %.

Appendix 4 (supplementary material) argues that the main results are not driven by unequal income growth across the groups.