Identifying Direct and Indirect Effects. Estimating the Costs of Motherhood Using Matching Estimators.

Marianne Simonsen*
Department of Economics
University of Aarhus

Lars Skipper
Department of Economics
University of Aarhus

Abstract

This paper discusses identification of direct effects when covariates are affected by the treatment. We seek to establish whether motherhood causes lower wages for women and estimate both net and direct effects of motherhood on wages implementing residual based matching and find small but significantly negative impacts. We extend the matching analysis to account for discrete covariates affected by the treatment, which allows us to identify direct effects under less strict assumptions. Applying this indicates major differences in impacts across public and private sectors, likely to be due to considerable job flexibility in the public sector.

JEL Classifications: C13, C14, C35, J13, J31, J45, J71

Keywords: Fertility, family gap, wages, matching.

Acknowledgements: We thank the Danish Social Science Research Council (the FREJA grant) for financial support and the National Centre for Register-based Research for support to access data from Statistics Denmark. We wish to thank Edward J. Vytlacil, Shelly Lundberg, Nabanita Datta Gupta, Michael Rosholm, Helena Skyt Nielsen, and Amalia R. Miller for their helpful suggestions. We also appreciate comments from participants at the EALE Conference 2003 and the Labor economics discussion group at Dept. of Economics, Stanford University. All errors are our own.

*Corresponding address: Marianne Simonsen, Dept. of Economics, University of Aarhus, Building 326, 8000 Aarhus C, Denmark. Email: msimonsen@econ.au.dk. Phone: +45 8942 1599. Fax: +45 8613 6334.

1 Introduction

The most common counterfactual of interest in the treatment evaluation literature is the mean direct effect of treatment on the treated (Heckman, LaLonde & Smith (1999)). In the case where the subpoulation under study consists of women, the treatment is being a mother, and the outcome is log wages this counterfactual has in the literature been designated 'the family gap' or, alternatively, 'the child penalty' (Budig & England (2001), Phipps, Burton & Lethbridge (2001), Waldfogel (1998*a*),Waldfogel (1998*b*)). The naming of the counterfactual is, of course, a direct consequence of the results found in the existing literature: Having a child seems to be costly in terms of wages.[1]

As noted in Waldfogel (1998*b*) we should be concerned with the family gap for two reasons: Equity and efficiency. Equal pay for equal work is obviously not achieved if discrimination is present. For one thing, this is important since the majority of women choose to have children at some point in their life and the family gap is a significant contributor to the gender wage gap (Waldfogel (1998*b*)). Furthermore, discrimination is likely to prevent the efficient utilisation of women's abilities which must be in the interest of the society as a whole. Thus it is important to understand what generates wage gaps between women with and without children.

Measuring the wage differences is not a simple task, however, since the choice of motherhood is likely to be endogenous. Angrist & Evans (1998) use exogenous variation in fertility in the form of preferences for sibling mix and the occurrence of twin births to estimate the effect of an extra child using exogenous variations in fertility.[2] The latter approach is also applied in Jacobsen, Wishart Pearce III & Rosenbloom (1999). Klepinger, Lundberg & Plotnick (1999) use instrumental variables analysis to account for the potential endogeneity of adolescent fertility. Also, motherhood may affect wages through several channels and hence there are potentially more parameters of interest. There may be a *direct* effect of motherhood on wages, i.e. the causal effect of motherhood on wages and there may be *indirect* effects running through the effect of motherhood on other covariates. (See Korenman & Neumark (1992) and Robins & Greenland (1992) for this terminology). For example, mothers may have lower levels of labour market experience due to child rearing activities and non-participation. This is likely to affect wages but the effect is *indirect*. We define the sum of the direct and the indirect effect as the net effect. The direct effect of motherhood on wages has, by far, received the most attention in the literature yet the recovering of the parameter is not immediate due to endogeneity bias introduced by conditioning on variables that are likely to be affected by the choice of motherhood, see discussion below. Both the direct and the net effect are interesting from a policy point of view: The former is needed if we want to make any conclusions on discrimination while the latter provides information on the costs of choices that are related to childbearing. Similar issues arise in the literature analysing the effects of college quality as emphasized in a recent paper by Black & Smith (Forthcoming). Here, years of schooling depend partly on college quality but also have a separate exogenous effect on labour market outcomes. The paper presents results from propensity score matching both with and without years of schooling included in the conditioning set and find large and significant differences in treatment effects.

One approach in the family gap literature, cf. e.g. Phipps et al. (2001) and Waldfogel (1998*b*) is to start out with a simple linear specification of the outcome equation using only a limited conditioning set and then keep adding covariates (essentially applying a specific to general approach) which resembles the idea of a direct versus indirect effect. Yet the papers tend to be less explicit about the cause of the effects. Furthermore, as pointed out by Korenman & Neumark (1992) and discussed in Section 3, adding covariates to the conditioning set is not problem free due to potential endogeneity bias.

Obviously, the sizes of the family gaps established in the literature depend on the quality and accessibility of the conditioning variables and the methods applied and, not surprisingly, the results vary substantially; Phipps et al. (2001) find a gap of 17.2% while Datta Gupta & Smith (2002) find no evidence of a signficant penalty.

Data quality is cardinal in all empirical work. The earlier work often included potential rather than actual work experience in the earnings equations.[3] This problem is essential when we want to study the effects of having children on women's wages, simply because women with children are more likely to interrupt their careers than women without children. See e.g. Mincer & Polachek (1974) or Mincer & Ofek (1982) for studies focusing on interruption effects as such. Waldfogel (1998*a*) finds that the negative effect of children is reduced when actual work experience is included rather than potential work experience. It is also found that access to job-protected maternal leave has substantial positive wage effects in both England and the US, offsetting some of the negative wage effect from having children. This is explained by the higher propensity of returning to the former employer after childbirth, hence diminishing depreciation of firm-specific human capital and retaining good job matches. In the Danish case with mandatory job-protected maternal leave schemes Datta Gupta & Smith (2002) find that, when they control for foregone human capital accumulation during periods out of the labour market, the birth of a child does lead to a temporary wage loss compared to childless women. However, this earnings effect vanishes around the age of 45 years. Budig & England (2001) estimate the family gap using US data. They find that when interruptions, lost experience, and part-time work are taken into account, the family gap is reduced considerably. Another potentially important effect neglected in earlier studies is human capital depreciation during interruptions. For Canada, Phipps et al. (2001) find that not only failure to acquire human capital but also the depreciation of human capital during leaves significantly reduce the penalty associated with ever having a child. Albrecht, Edin, Sundström & Vroman (1999) investigate earnings effects from career interruptions by use of a Swedish data set and find evidence of depreciation during interruptions. The same results are found in the Danish case by Nielsen, Simonsen & Verner (Forthcoming).

It is striking that even after controlling for a vast number of conditioning variables there still exists a wage gap between mothers and women without children. There may be real productivity differences between the two groups of women, i.e. there may be omitted variables bias. Thus improving data quality is still important.

Common to all the analyses mentioned above is the assumption of a common treatment effect at least within subpopulations. As noted in Heckman et al. (1999) this is a very strong assumption. Furthermore, all are subject to assumptions on functional form of the equations of interest; for one thing separability of the effects of observables and unobservables is assumed, and the conditional expectations function is assumed to be linear in attributes. Some are also subject to parametric assumptions on correlations of error terms. Alternatively, one may apply matching analysis. Matching is based on the principle that comparing the outcome for an individual from the treatment group, i.e. mothers, with the outcome for an individual from the no treatment group, i.e. women without children, who in terms of observables is sufficiently similar to the treated individual on average, balances the selection bias arising from self selection into motherhood. Matching allows for heterogenous treatment effects, is not subject to parametric assumptions, and does not *per se* assume separability of the effects of observables and unobservables. It does, however, require extremely rich data sets.

In this paper we apply propensity score matching to estimate the effect of having children on women's wages. The data at our disposal is a Danish register based data set that includes very detailed, high quality information on e.g. income, demographics, and education on a yearly basis. Furthermore, the individual event history in terms of periods of employment, unemployment, maternal leave, and publicly subsidised leave (child rearing and sabbatical) is known on a weekly basis. We discuss the assumptions needed for identification of direct and net effects of being a mother when variables are affected by the treatment. In general, conditioning on covariates affected by the treatment is never justified (Rosenbaum (1984)), however under the assumption of linearity of the outcome equations this problem may be overcome. We take advantage of an exclusion restriction to perform regression adjusted matching recently suggested by Heckman, Ichimura & Todd (1998). To our knowledge matching analysis has never been applied in this area before and we know of very few examples of the implementation of regression adjusted matching in general (e.g. Smith & Todd (Forthcoming)). Furthermore, we extend the matching analysis to take account of discrete covariates that are affected by the treatment *without* assuming linearity of the outcome equation and apply this to sector choice. The public sector is characterised by having more flexible working conditions and both duration of and compensation during maternity leave are higher in the public sector. Hence, we expect there to be more mothers working in the public sector all other things being equal, cf. Nielsen et al. (Forthcoming). We find significant family gaps no matter the choice of matching estimator but the results vary between public and private employees. The direct effect is small indicating that most of the difference in wages can be explained by variables that are likely to be affected by having children.

The paper is organised as follows. In the next section we discuss our sample of data. In section 3 we discuss our parameter of interest and outline our econometric strategy. Section 4 presents the results and section 5 concludes.

2 Data

The original data set contains information on a representative sample of 5% of all Danish individuals in the 15-74-age bracket. Information stems from several registers all maintained by Statistics Denmark. The registers include variables describing income, demographics, and education on a yearly basis. Furthermore, the individual event history in terms of periods of employment, unemployment, maternal leave, publicly subsidised leave (child rearing or sabbatical), and the residual category non-participation is known on a weekly basis.

In the empirical analysis below, we use a 1997 cross sectional subsample of women aged 20-40 years who are employed more than 200 hours per year, who are not self-employed, and not undertaking education. The lower age bound is chosen to exclude individuals who are in the state between two types of education, for instance high school and university. The upper age bound is chosen because of an age restriction on the availability of parental information, used to construct the exclusion restriction applied in the econometric analysis. The analysis is performed using retrospective information on the labour market history.

Table 1 shows descriptive statistics on selected variables for the sample used in our analysis along with descriptive statistics of mothers and non-mothers. We classify women as mothers if they have given birth to a child. Thus it is assumed that the presence of biological children is more important than the presence of stepchildren. It may be a problem if children in the household other than biological children of the woman affect her choices and actions.[4] 54.8% of the women in our sample have children in 1996.

The outcome variable of interest in the analysis is log hourly wage. It is calculated from annual earnings and number of working hours. The measure of working hours used in this calculation is very precise in that the hours information comes from registers on compulsory contributions to supplemental pension payments that are closely linked to the working hours actually paid for by employers. It is seen that average log wage for mothers does not differ from average log wage for non-mothers. Yet it is clear from table 1 that mothers differ significantly from non-mothers in terms of observables: Mothers are on average 7 years older, they have twice as much labour market experience, are more likely to have an education directed towards the health care sector or the schooling system, are more often employed in the public sector, and have on average longer unemployment spells. Furthermore, they are more likely to own real estate and are more often settled in the province. Finally, they are, not surprisingly, more likely to be married and have fewer siblings of their own. This latter variable will be used as our exclusion restriction in what follows. We argue below that number of siblings is correlated with fertility but not with productivity, i.e. labour market outcome.

[**Table 1 around here**]

The information on interruptions consists of a subset of spells created from the accurate event histories known on a weekly basis. Incidences of unemployment and non-participation are registered from 1981 and onwards while maternity leave and parental leave in connection with childbirth can be traced back to 1984. Before 1984, maternity leave is included in the non-participation category. In the period before 1984, mothers are eligible for 18 weeks of maternity leave (4 weeks before expected birth and 14 weeks after), which means that for the oldest women in our sample some maternity leave may be hidden in the residual category of non-participation. In 1984, the maternity leave scheme is extended with 10 weeks of leave amounting to a maximum of 28 weeks and fathers are granted 2 weeks of leave during the first 14 weeks after the birth. This is the scheme effective in 1997. The 10 additional weeks can in principle be shared with the father, yet this is not the norm. In fact, in 1997 96.6% of all fathers on leave take 2 weeks or less and the mode is 2 weeks (cf. Statistics Denmark - www). The degree of compensation while on maternity leave varies with union membership but with a legally ensured lower bound on benefits received.

In 1994, publicly subsidised sabbatical leave and child rearing leave were introduced. For employed individuals child rearing leave amounts to a maximum of 52 weeks per child under the age of 8, while sabbatical leave amounts to a maximum of 52 weeks. The length of these 2 types of leaves is registered from 1995 and onwards. For estimation purposes we choose to pool non-participation with maternity leave, parental leave, sabbatical, and child rearing leave.

3 The Parameter of Interest

In this section we will briefly discuss the objectives of the econometric analysis and notation and continue on to discuss our parameter of interest.

The goal of the evaluation is to measure the effect or impact of a given treatment, C, on an outcome variable, Y. Here the treatment is having children and the outcome of interest is wage growth. Let Y_{1i} be the person-specific outcome in the presence of children, and

Y_{0i} the outcome in the absence of children. Hence, the person-specific impact of having children is defined as $\Delta_i \equiv Y_{1i} - Y_{0i}$. The Fundamental Evaluation Problem is that we do not observe both Y_1 and Y_0 for anyone at the same point in time. This also applies in the case under study: Never do we observe the same woman both with and without children at the same point in time. It therefore becomes impossible to construct the person-specific impact for any woman by simply looking at the data. Instead, the interest usually shifts in the literature from constructing personal impacts to constructing conditional means.

The parameter most often estimated in the literature is the mean direct effect of treatment on the treated defined as

$$\begin{aligned} \theta &\equiv E\left[Y_1 - Y_0 | C = 1\right] \\ &= E\left[Y_1 | C = 1\right] - E\left[Y_0 | C = 1\right]. \end{aligned} \tag{1}$$

Hence, the problem turns to that of finding the counterfactual $E\left[Y_0 | C = 1\right]$ in (1), which is, of course, unobserved. I.e., some assumptions are needed to obtain identification.

Much applied microeconometric analysis is concerned with measuring the degree of some kind of discrimination, e.g. discrimination of women or ethnic minorities. In focusing on potential discrimination of women on the labour market with respect to wages it is found in e.g. Waldfogel (1998*b*) that the most important factor in explaining gender wage gaps is differences in returns to family related characteristics between genders. It is consistently found that women are punished in terms of wages when having a child as opposed to men who are found to experience higher wage growth (Jacobsen & Rayak (1996), Korenman & Neumark (1991), Loh (1996)). As noted in Waldfogel (1998*b*) we should be concerned with the family gap for two reasons: Equity and efficiency. Equal pay for equal work is obviously not achieved if discrimination is present. Furthermore, discrimination is likely to prevent the efficient utilisation of women's abilities which must be in the interest of the society as a whole. Thus it is important to understand what generates wage gaps between women with and without children and the first step in doing that must be to precisely measure the gap and establish whether the size is robust or just an artefact of rather restrictive estimation procedures. Obviously, the effect of the treatment 'having children' cannot be evaluated using social experiments and we are as all others forced to use observational estimators.

We estimate the average effect of having children on those women who choose to have children, ATET, (1), and not the average treatment effect, ATE. The former is interesting from a policy point of view: It is interesting whether women who choose to have children are punished in terms of wages, not so much whether women who do not have children *would* be punished had they chosen to have children. The estimated parameter is, of course, not structural and may as all non-structurally estimated parameters be subject to the Lucas critique. However, as mentioned above it is still an economically interesting parameter: If all analyses consistently find significant family gaps even when controlling for a comprehensive set of conditioning variables applying sophisticated estimation methods there does indeed seem to be a real problem that needs looking into.

In this paper we choose to apply the method of matching, that has recently received much attention in applied econometrics in general and in programme evaluation in particular. Matching is based on the assumption that conditioning on observables, $\mathbf{X}$, eliminates the selective differences between those with and without children. More precisely, the method of matching assumes that the econometrician has access to conditioning variables sufficiently rich such that the counterfactual outcome distribution of those having children is the same as the observed outcome distribution of those without. By conditioning on the covariates at our disposal, we will thus be capable of balancing the bias coming from the self-selection into motherhood.

In focusing on (1) we make the following assumption[5]

$$E\left[Y_0\left|\mathbf{X}, C=1\right.\right] = E\left[Y_0\left|\mathbf{X}, C=0\right.\right] = E\left[Y_0\left|\mathbf{X}\right.\right]. \qquad (\mathbf{A\text{-}1})$$

In other words, we assume that the expected outcome for a mother *had she not had children* is the same as the expected outcome for a non-mother with the same observed characteristics. In particular, this means that women must not take into account wages in the non-motherhood state when deciding whether to become mothers or not. They may, however, consider wages in the motherhood state. This is consistent with the case where mothers and non-mothers are equally productive in the non-motherhood state conditional on their characteristics but are potentially different in the motherhood state. In order to be able to utilise (**A-1**) it is necessary to make sure that there is a woman without a child analogue to each mother, i.e.,

$$P \equiv \Pr\left(C=1\left|\mathbf{X}\right.\right) < 1 \qquad (\mathbf{A\text{-}2})$$

Notice that we do *not* assume away the selection bias. Instead we simply rebalance the bias.

At this point we do not want to assume any functional form of the outcome equation as opposed to much of the literature already mentioned in the introduction. We are therefore potentially faced with the nonparametric curse of dimensionality due to our rich register data. A way to circumvent the curse of dimensionality without imposing arbitrary assumptions is based on the results in Rosenbaum & Rubin (1983). Here the focus is shifted from the set of covariates to the probability of motherhood, $P_i = \Pr(C = 1 | \mathbf{X}_i)$. As long as (**A-1**) and (**A-2**) hold,

$$E[Y_0 | P, C = 1] = E[Y_0 | P, C = 0] \qquad (\mathbf{A\text{-}3})$$

over the common support.[6] This new conditioning variable, P, changes our conditional mean assumption into (**A-3**) which together with $P < 1$ are the only conditions required in order to justify propensity score matching to estimate the mean impact on the treated, see Heckman, Ichimura & Todd (1998). In appendix A we discuss the details in implementing the actual matching procedure chosen.

We also perform regression adjusted matching as recently suggested in Heckman, Ichimura & Todd (1997) and Heckman, Ichimura & Todd (1998). An additional necessary assumption is additive separability of observables and unobservables in the outcome equations; an assumption usually invoked in conventional econometric models. To be able to identify this model without restrictions on functional form of the conditional expectation function of the error term we need an exclusion restriction. We utilise the parental information mentioned above. Formally, we partition our explanatory variables into two sets, $\mathbf{X}$ and $\mathbf{X}_c$, where $\mathbf{X}_c$ contains at least one variable not in $\mathbf{X}$. Specifically, we assume

$$Y_0 = \mathbf{X}\boldsymbol{\beta}_0 + U_0. \qquad (2)$$

Furthermore, **A-3** is replaced by

$$E[U_0 | P(X_c), C = 1] = E[U_0 | P(X_c), C = 0]. \qquad (\text{A-3'})$$

To estimate $\boldsymbol{\beta}_0$ we apply partially linear regression as in Heckman et al. (1997), see appendix A for technical details. We clear out the effects of potential returns in the non-mother state to observable characteristics from Y and perform matching on the residuals.

Obviously, residual based matching is based on stronger assumptions that the non-residual based version and in the case where (2) is the true specification the estimated average impacts are asymptotically equivalent. However, this statement implicitly presupposes exogeneity of the variables included in X. In the next section we seek to clarify the effects on the estimated parameters due to endogenous variables in X.

3.1 Identification of the parameter of interest when information in the conditioning set is determined by the treatment

A potential problem in our set-up, as in many other applications, e.g. Black & Smith (Forthcoming), is endogeneity of regressors in the main equation. In particular, we expect neither experience nor the interruption variables to be exogenous wrt. fertility: Presumably, mothers are more likely to interrupt their careers to engage in child-rearing activities, i.e. non-participation or formal child-rearing leave. Rosenbaum (1984) examines the consequences for the average treatment effect of including a variable, S, that has potentially been affected by the treatment in the conditioning set. He finds that, in general, adjustment for such a variable results in unbiased estimates of the average treatment effect only when the variable is in fact not affected by the treatment. Furthermore, existence of a variable that has been affected by the treatment introduces an indirect effect of the treatment. Robins & Greenland (1992) discuss the difficulties in identifying the direct effect of the treatment when such a variable is present. To illustrate the consequences in the case where the true model is linear, consider the following simple setup:

$$\begin{aligned} Y_0 &= \alpha + \beta_1 S_0 + U \\ Y_1 &= \alpha + \beta_0 + \beta_1 S_1 + U \\ S_0 &= \lambda + \gamma_1 Z + V \\ S_1 &= \lambda + \gamma_0 + \gamma_1 Z + V \\ C &= C\left(X_c, \varepsilon\right) \end{aligned} \tag{3}$$

where subscript 0 refers to the no-treatment state, $C = 0$, and subscript 1 refers to the treatment state $C = 1$. The outcome equation Y is assumed to be linear in the regressors and, for simplicity, we assume a constant *direct* effect of being a mother, β_0. S is a variable affecting outcome that is potentially affected by the treatment, C. Being af mother is some function, $C(\cdot)$, of X_c and an error term ε.

The semi-reduced form is

$$Y = \alpha + (\beta_0 + \beta_1\gamma_0)C\left(X_c, \varepsilon\right) + \beta_1\gamma_1 Z + (\beta_1 V + U)$$

Estimation is based on the expected attribute adjusted treatment difference (Rosenbaum (1984)):

$$E_X\left[E\left[Y_1|C=1,X\right] - E\left[Y_0|C=0,X\right]|C=1\right]$$

where X is some conditioning set.

Firstly, consider the case where we condition on X_c alone. The conditional expected difference in outcome is:

$$\begin{aligned} E\left[Y_1|C=1,X_c\right] - E\left[Y_0|C=0,X_c\right] &= \beta_0 + \beta_1\gamma_0 + \\ &\quad \beta_1\gamma_1\left\{E\left[Z|C=1,X_c\right] - E\left[Z|C=0,X_c\right]\right\} + \\ &\quad E\left[\beta_1 - U|C=1,X_c\right] - E\left[\beta_1 V - U|C=0,X_c\right]. \end{aligned}$$

We interpret $\beta_0 + \beta_1\gamma_0$ as the net effect of being a mother on wages in this special case. The term $\beta_1\gamma_1\left\{E\left[Z|C=1,X_c\right] - E\left[Z|C=0,X_c\right]\right\}$ results from potential differences in the distribution of Z conditional on X_c among mothers and non-mothers. Identification of the net effect requires that $(Z,\ \beta_1 V - U)$ is mean independent of C conditional on X_c: $E\left[Z|C=1,X_c\right] = E\left[Z|C=0,X_c\right]$ and $E\left[\beta_1 V - U|C=1,X_c\right] = E\left[\beta_1 V - U|C=0,X_c\right]$.

Secondly, consider the case where we condition on both X_c and Z:

$$\begin{aligned} E\left[Y_1|C=1,Z,X_c\right] - E\left[Y_0|C=0,Z,X_c\right] &= \beta_0 + \beta_1\gamma_0 + \\ &\quad E\left[\beta_1 V - U|C=1,Z,X_c\right] - \\ &\quad E\left[\beta_1 V - U|C=0,Z,X_c\right]. \end{aligned}$$

In this case we only need $E\left[\beta_1 V - U|C=1,Z,X_c\right] = E\left[\beta_1 V - U|C=0,Z,X_c\right]$ to identify the net effect of having a child.

Thirdly, consider the case where the variable that is potentially affected by the treatment is included in the conditioning set:

$$
\begin{aligned}
E\left[Y_1|C=1,S_1=s,X_c\right]-E\left[Y_0|C=0,S_0=s,X_c\right] &= \beta_0+ \\
&\quad E\left[U|C=1,S_1=s,X_c\right]- \\
&\quad E\left[U|C=0,S_0=s,X_c\right].
\end{aligned}
$$

In this case we can identify the *direct* effect from having a child if $E\left[U|C=1,S_1=s,X_c\right]= E\left[u|C=0,S_0=s,X_c\right]$. In particular, this holds if $E\left[U|C=1,S_1=s,X_c\right]=E\left[U|C=1,X_c\right]= E\left[U|C=0,S_0=s,X_c\right]=E\left[U|C=0,X_c\right]$.

Finally, consider the case where we condition on the full set of information:

$$
\begin{aligned}
E\left[Y_1|C=1,S_1=s,Z,X_c\right]-E\left[Y_0|C=0,S_0=s,Z,X_c\right] &= \beta_0+ \\
&\quad E\left[U|C=1,S_1=s,Z,X_c\right]- \\
&\quad E\left[U|C=0,S_0=s,Z,X_c\right].
\end{aligned}
$$

Even in this case we must have that the expectation of the error term in the outcome equation is conditionally mean independent of S to avoid bias.

In conclusion, if we do not want to restrict the functional form of the outcome equation, we can only condition on variables that are not affected by the treatment and hence just hope to estimate the net effect of having children. Furthermore, this requires some assumptions on the distribution of variables Z affecting S as illustrated by the linear example. Imposing a linear structure on the outcome equation allows us to identify the direct effect of having children if we are willing to restrict the error term in the outcome equation to be conditionally mean independent of C and S.

3.2 Discrete attributes determined by the treatment

The section above illustrates the difficulties arising when outcome is affected by variables that are determined by treatment. If attributes affect wage determination and are determined by the treatment ignoring the information leads to non-identification of the direct effect. The net effect of treatment may, however, be estimated under assumptions on the distribution of the variables determining S so one alternative strategy would be to redefine the parameter of interest. Estimating the direct effect requires fairly strict assumptions; in particular, there may be problems if the outcome equations are non-linear. This section describes a way to take the endogenous nature of an attribute, S, into account in the case where it takes on discrete values. For simplicity, let S be binary. The following is easily extended to include cases where S takes on a arbitrary (finite) number of discrete values. Then we have two types of simultaneous treatments, C and S, and four mutually exclusive states:

	$C=0$	$C=1$
$S=0$	Y_{00}, P_{00}	Y_{10}, P_{10}
$S=1$	Y_{01}, P_{01}	Y_{11}, P_{11}

where P_{ij} is the probability of being in state ij. The parameters of interest are different average treatment effects (suppressing the conditioning set):

The unconditional average treatment effect

$$E\left(Y_{ij}-Y_{lm}\right),$$

the average treatment effect conditional on both treatments

$$E\left(Y_{ij}-Y_{lm}|C=i,S=j\right),$$

the average treatment effect conditional on treatment X

$$E(Y_{ij}-Y_{lm}|S=j),$$

and the average treatment effect conditional on treatment C

$$E(Y_{ij}-Y_{lm}|C=i).$$

Note that both $E(Y_{ij}-Y_{lm}|S=j)$ and $E(Y_{ij}-Y_{lm}|C=i)$ are weighted combinations of versions of $E\left(Y_{ij}-Y_{lm}|C=i,S=j\right)$:

$$\begin{aligned} E(Y_{ij}-Y_{lm}|S &= j)=E(Y_{ij}-Y_{lm}|C=1,S=j)P(C=1|S=j) \\ +E(Y_{ij}-Y_{lm}|C &= 0,S=j)P(C=0|S=j) \end{aligned}$$

$$\begin{aligned} E(Y_{ij}-Y_{lm}|C &= i)=E(Y_{ij}-Y_{lm}|C=i,S=1)P(S=1|C=i) \\ +E(Y_{ij}-Y_{lm}|C &= i,S=0)P(S=0|C=i). \end{aligned}$$

Thus identifying $E(Y_{ij} - Y_{lm}|S = j)$ and $E(Y_{ij} - Y_{lm}|C = i)$ is a question of identifying $E(Y_{ij} - Y_{lm}|C = i, S = j)$ and the conditional probabilities.

The proof of the generalisation of the balancing score property in our case with simultaneous treatments follows the literature on multiple treatments (Lechner (2001) and Imbens (2000)). Define T

$$T = \begin{cases} 0, \text{ if } C = 0, S = 0 \\ 1, \text{ if } C = 0, S = 1 \\ 2, \text{ if } C = 1, S = 0 \\ 3, \text{ if } C = 1, S = 1 \end{cases}.$$

Then the result from Lechner (2001)

$$\begin{aligned} Y_{00}, Y_{01}, Y_{10}, Y_{11} \amalg T|Z &= z \Rightarrow \\ Y_{00}, Y_{01}, Y_{10}, Y_{11} \amalg T|b(Z) &= b(z), \quad \forall z \in \chi, \\ \text{if } \; E\left[P(T = q|Z = z)|b(Z) = b(z)\right] &= P\left[T = q|Z = z\right], \end{aligned}$$

$$0 < P\left[T = q|Z = z\right] < 1, \; \forall q = 0, , , Q,$$

where χ is the attribute space, holds true in our case as well. The common support requirement in the standard matching literature invoked to obtain identification is then extended to $P^q(z)P^r(z) > 0$. In our case this corresponds to identification of the average unconditional treatment effect, $E(Y_{ij} - Y_{lm})$, and the average treatment effect on the treated, $E(Y_{ij} - Y_{lm}|C = i, S = j)$ as long as $P(C = i, S = j|Z = z)P(C = l, S = m|Z = z) > 0$.

The estimation procedure deviates from Lechner (2001) on a very important point: The probability model $P(\cdot)$ must take into account that S is potentially affected by C. Also, we do not want to restrict the correlation between unobserved characteristics affecting C and S. Hence an appropriate (parametric) model could be a bivariate probit.

To illustrate the point of this exercise reconsider the linear example in (3) in the section above with the only difference that S is binary. This corresponds to the following model:

$$
\begin{aligned}
Y_{00} &= u \\
Y_{01} &= \beta_1 + u \\
Y_{10} &= \beta_0 + u \\
Y_{11} &= \beta_0 + \beta_1 + u \\
S &= S(C, Z, e) \\
C &= C\left(X_c, \varepsilon\right)
\end{aligned}
$$

Note that we can identify the direct effect of having children *without* assuming that u is mean independent of S:

$$
\begin{aligned}
E\left[Y_{10}|C=1, S=0, X\right] - E\left[Y_{00}|C=0, S=0, X\right] &= \beta_0 + \\
&\quad E\left[u|C=1, S=0, X\right] - \\
&\quad E\left[u|C=0, S=0, X\right]
\end{aligned}
$$

where X is the conditioning set.

4 Results

We consider a generalisation of the model presented in (3): Outcome, Y, log hourly wages, is some function $F(\cdot)$ of whether the woman is a mother or not, $C=1$ or $C=0$, observed characteristics, $\mathbf{X}$, and unobserved characteristics, u.

$$Y = F(C, X, u)$$

A subset of the observed characteristics $\mathbf{S} \subseteq \mathbf{X}$ is potentially affected by the treatment, that is whether the woman is a mother or a non-mother. $S_j \in \mathbf{S}$ is some function $G_j(\cdot)$ of C, observed characteristics $\mathbf{Z}_j$, and unobserved characteristics, v_j.

$$S_j = G_j(C, Z_j, v_j)$$

Finally, the probability of being a mother, the propensity score, is a function C of observed characteristics $\mathbf{X}_c$ and unobserved characteristics ε. A subset of the variables in $\mathbf{X}_c$ affect outcome Y.

$$C = C(X_c, \varepsilon)$$

In the following we characterise selection into motherhood and present results from the matching analyses.

4.1 Estimation of the Propensity Score

The first step in the analysis is to estimate the probability of being a mother. We model the propensity score by a standard probit.[7] As mentioned above Rosenbaum (1984) shows that matching on variables that are affected by the treatment will in general give biased estimates. Furthermore, even if we assume partial linearity of the outcome equation, the direct effect of being a mother can only be identified if $E[U|C=1, S_1=s, X_c] = E[U|C=0, S_0=s, X_c]$, i.e. U is conditionally mean independent of S as discussed in section.3.1. Hence, we seek to avoid including variables in the propensity score that we suspect to be affected by the treatment. We condition on age, type of education e.g. health care, services, technical education etc., length of education given type, place of habitation outside capital area (greater Copenhagen), and the woman's number of siblings.[8] The results are presented in table 2.

We find that age significantly increases the probability of being a mother in 1996 with a decreasing effect reaching its maximum at the age of 44, which is outside the range of our data. Hence, the effect is increasing over the relevant range. The variable is, of course, deterministic and thus by no means endogenous to the decision of motherhood. Furthermore, relative to health care oriented types of education most types of education decrease the probability of being a mother for a given length of education, though some coefficients are insignificant. Note that some may be insignificant due to a small number of individuals in a particular category - see table 1. The education variables presumably capture some effects of preferences for having children. The analysis also shows that higher level of education of a given type reduces the probability of motherhood. The choice of education is in the vast majority of cases predetermined to the decision of motherhood but we acknowledge that unobserved factors may affect both the choice of education and fertility. As long as (**A-1**) holds this is not a problem. Habitation outside the capital area also increases the probability of motherhood. This variable is meant to capture the effects of a possibly more traditional view on having children. Finally, the woman's number of siblings significantly increases the probability of having children. This variable is our exclusion restriction when performing residual based matching. It may on all reasonable grounds be left out of the wage function; we do not expect number of siblings to affect labour market productivity conditional on the information in X described in the next section and it is definitely predetermined to wage determination. Furthermore, the variable is correlated with the choice of having children since it enters the selection equation with a significant coefficient. Butcher & Case (1994) and Ermisch & Francesconi (2001) analyse the effect of family composition on wages and they both argue that number of siblings affect wage determination only through the individual's level of education, which supports our use of the variable as an exclusion restriction: We include the woman's type and length of education in the wage function and conditional on that information, the woman's own number of siblings should not be correlated with the error term, U, in the outcome equation.

[Table 2 around here]

[Figure 1 around here]

The model predicts relatively well: 5867 out of 29210 predictions or 20.01% of all predictions are wrong[9] and Efron's R^2 equals 0.42.[10] Figure 1 shows the smoothed densities of the propensity scores for both mothers and non-mothers. It is seen that mothers have more probability mass concentrated around high values of the propensity score compared to non-mothers who have more mass concentrated around low values of the propensity score. Hence, mothers are likely to differ significantly from non-mothers in terms of observables meaning that there is a potential gain from matching. Note that the densities seem to have common support: It is possible to find a match for a mother among non-mothers even for mothers with the highest level of the propensity score. Thus the model does not predict too well. We also conclude that we obtain reasonably balanced covariates after matching on our estimated propensity score.[11] See Appendix B for details.

4.2 Propensity Score Matching Analysis

As indicated in section 2, in our sample mothers differ significantly from non-mothers in terms of observables. Comparing outcome for mothers with outcome for non-mothers without conditioning on observables is therefore problematic; we expect the positive difference in raw means to be attributable to differences in e.g. age and hence experience between the two groups.

The goal of our estimations is twofold: We wish to evaluate both the average net and the average direct effect of being a mother. We interpret the direct effect as the causal effect of motherhood on wages. This may include both discrimination, statistical discrimination, and bargaining differences. The net effect includes the direct effect as well as indirect effects resulting from mothers' potentially lower labour market experience, higher level of non-participation, and greater probability of working in the public sector which provides better working conditions for families, to mention a few examples. The net effect is an estimate of the total wage cost of having children whereas the direct effect is the causal effect of being a mother on wages.

Firstly, we perform non-residual based matching: We compare log wages for mothers with log wages for non-mothers that are similar in terms of the propensity without making any assumptions on the functional form of the outcome equation, F. This corresponds to the following expected attribute adjusted treatment difference:

$$E_{X_c}\left[E\left[Y_1|C=1,X_c\right]-E\left[Y_0|C=0,X_c\right]|C=1\right]$$

This gives us an estimate of the net effect of being a mother on wages including *all* effects stemming from S being potentially affected by the treatment. Without a formal economic model of the determinants, Z, of the variables in S, we do not want make the assertion that their distributions do not vary among mothers and non-mothers. We define the set S to consist of level of experience, past unemployment and non-participation information, level of occupation, and choice of sector. The identifying assumptions are the usual assumptions in the matching literature: (**A-1**) and (**A-2**). In other words, we must have enough information in X_c to exactly balance the bias. We see from table 3 that this estimator results in a 6.6% lower wage growth for mothers; a result in the lower end of the estimates in the international literature.

We also perform regression adjusted matching imposing (2) conditioning on the same variables as when performing non-residual based matching and the exclusion restriction. In other words X and X_c include the same variables except for the exclusion restriction in X_c. We find a comparable penalty of 7.0% lower wage growth for mothers. A large difference in these results would tend to suggest that the linear structure of the outcome equation was too restrictive. Obviously, the statistical insignificant difference in estimates does not *prove* that the partial linear structure is correct!

If we assume additive separability of the outcome euqations we can perform regression adjusted matching and clear out the effect of the variables potentially affected by the treatment by including them in X in (2). To do this we need consistent estimates of β, which requires $U \amalg S|X_c$ (cf. Robinson (1988)), a strong condition that implies conditional mean independence discussed in section 3.1. Under the assumptions given, this identifies the *direct* effect of being a mother on wage growth. The variables in X_c are the same as in the preceding estimation. This estimator is based on the following expected attribute adjusted treatment difference:

$$E_{X,X_c}\left[E\left[Y_1|C=1,X,X_c\right]-E\left[Y_0|C=0,X,X_c\right]\right].$$

When specifying the conditioning variables in the log wage function, X, we follow traditional human capital theory.[12] Apart from the variables included in standard wage equations such as actual work experience, actual work experience squared, very specific types of education, length of these types of education, choice of sector, and occupation categories we include a series of other variables to allow for various effects of interruptions (cf. Nielsen et al. (Forthcoming)). Figure 2 illustrates the potential effects of interruptions on the earnings potential and emphasises the link between the theoretical effects and our explanatory variables. We consider a woman who interrupts her career to have a child engaging in child-rearing activities between age A0 and age A1. During the interruption, the woman fails to accumulate experience. This simply corresponds to a horizontal shift in the earnings profile. This effect is caught in our model by using actual experience as the conditioning variable. In addition, earnings profiles may shift downwards because human capital depreciates while interrupting with the effect possibly depending on the duration of the interruption. We allow for linear depreciation as illustrated in the figure and condition on the total duration of the interruption spells of a given type. Finally, there may be a catching up effect: Women may regain part of (or all) the effect of lost experience and depreciation when they return to work. In the literature, the period of catching up has been labelled "the recovery phase". We include yearly indicators for the timing of the end of the last interruption spell thus allowing for the effect of the interruption to decrease in time. The types of interruptions considered are unemployment and non-participation. Since regression adjusted matching clears out the effects of potential returns in the non-mother state we are forced to pool the non-participation category with maternity leave and other publicly subsidised types of leave to be able to account for the effects of these types of interruption We do not expect this assumption to be too restrictive; there are no ex ante reason why depreciation of human capital during non-participation should differ from that of maternity leave.

We find a much smaller penalty of 1.5% when conditioning on this additional information. In fact, this corresponds to a reduction of almost 80% compared to regression adjusted matching excluding this information. The identifying assumptions needed to recover the direct effect of being a mother are strong! It must, for example, be the case that the expected unobserved characteristic for a mother is the same as for a non-mother conditional on them having the same X_c even though the distribution of experience may differ among the two groups. Note also that these assumptions - *along* with an assumption of constant treatment effect - are commonly made in the existing literature. In Appendix D we discuss sensitivity of our results and check the internal consistency of the model.

[Table 3 around here]

[Figure 2 around here]

4.3 Discrete Attributes affected by the treatment: The Case of Sector choice

As mentioned above we can avoid having to assume linearity of the outcome equations if the covariate affected by the treatment only takes on discrete values and still identify the direct effect of having children. Obviously, we have more than one covariate that is potentially affected by the treatment, but at the very least we can avoid assuming that discrete attributes enter linearly in the outcome equations. One variable of particular interest is sector choice: The public sector is characterised by having more flexible working conditions and both duration of and compensation during maternity leave are higher in the public sector. Hence, we expect there to be more mothers working in the public sector all other things being equal, cf. Nielsen et al. (Forthcoming). Table 4 shows the distribution of mothers in the two sectors.

[Table 4 around here]

In this case with simultaneous treatments there is more than one parameter of interest as described in Section 3.2. The first parameter of interest, suppressing the conditioning set X, is

$$E\left(Y_{10} - Y_{00} | C = 1, S = 0\right),$$

that is the expected difference in log hourly wages between mothers and non-mothers in the public sector conditional on motherhood and public sector, while the second is

$$E\left(Y_{11} - Y_{01} | C = 1, S = 1\right),$$

the expected difference in log hourly wages between mothers and non-mothers in the private sector conditional on motherhood and private sector. We estimate both net and direct effects of motherhood as in the section above.

Firstly, we perform matching for the two sectors separately, i.e. *without* taking the potential endogeneity of sector choice into account. The results from estimating ATET using non-residual based matching compare reasonably well to the results from the full sample, see Table 5 below.[13] We find that mothers in the public sector have a significant 6.5% lower wage growth while mothers in the private sector have a smaller significant loss of 5.3% in terms of wage growth. The difference is not significant, though. The results from residual based matching, on the other hand, differ very much from the results from the full sample: Mothers in the public sector have a significant 3.5% lower wage growth while mothers in the private sector are not punished in terms of wages. The point is, however, that sector choice is not random! For example, it may be that mothers in the public sector would have had much lower wage growth had they worked in the private sector.

[Table 5 around here]

Next, we use the method suggested in Section 3.2 to allow for the potential endogeneity of sector choice. As in the section above we start out with estimation of the propensities.

We use a bivariate probit model to estimate the probabilities of being a mother and working in the private sector allowing the error terms in the two equations to be correlated. From this model we can get the predicted simultaneous and conditional propensities we need to identify the parameters of interest. The selection into motherhood is modelled using the same variables as before including the information on number of siblings.

[Table 6 around here]

When modelling the sector selection we condition on motherhood indicator, type of education, length of education given type, and place of habitation outside capital area (greater Copenhagen). Furthermore, we include parental information on sector choice. The results from the bivariate probit can be seen in Table 6.[14] It is obvious from Table 2 and 6 that the coefficients in the motherhood equation have not changed substantially by allowing the error term to be correlated with sector choice. The coefficients from the sector choice equation indicate that mothers are less likely to choose to work in the public sector in line with with our hypothesis, most types of education increase the probability of working in the private sector compared to health care related types of education, whereas the effect of length of education vary.

The information on number of siblings and the parental information on sector choice serve as exclusion restrictions. Importantly, the exclusion restriction remain significant and the coeffecient seems robust to the change of model. We argue that number of siblings is uncorrelated with both wage outcome (cf. discussion above) and sector choice and that parental sector choice is uncorrelated with wage outcome and fertility. Both arguments rely on conditioning on the information in X_c, $Z_{\sec tor}$ and X, where $Z_{\sec tor}$ is the information determining sector choice. Their inclusion serves two purposes: Firstly, we avoid that identification in the motherhood-sector choice model relies too heavily on the joint normality of the error terms and secondly, it allows us to perform residual based matching.

[Table 7 around here]

The results from the matching analysis taking endogeneity of sector choice into account can be seen in Table 7. It is clear that accounting for the endogeneity of sector choice matters. Note that we cannot use the estimates of the net effects to conclude that being a mother is much more expensive in the public sector. The difference may to a large degree reflect that mothers in the public sector make different choices than mothers in the private sector in terms of non-participation, part time, choice of occupation etc. In principle, in the public sector, wages for mothers that have been on maternity or child rearing leave should not differ from non-mothers wages everything else being equal, since wages are highly correlated with seniority in particular in the period under consideration. However, being on leave may affect a mother's chances of getting a socalled 'qualification bonus' given in the public sector. Along with promotion opportunities this is the mean managers in the public sector can use to reward non-mothers for not taking leave. Since it is not related to level of experience, non-participation etc. we must expect it to show up in the direct effect. The penalty could also be caused by job flexibility, a nonpecuniary benefit, within the public sector: Besides having extra care days targeted towards childrens needs mothers may be reallocated to less stressful or timeconsuming jobs, for example jobs involving only standard hours or less meeting activity.

It is noteworthy that the direct effect of motherhood is small and positive in the private sector. One explanation for the size of the direct wage effect in the private sector may be that wages in the private sector are much more flexible such that wage outcome may better reflect differences in observables. The fact that we find a wage *premium* in the private sector may be somewhat surprising. For us to be able to interpret the parameter as a direct effect it must be the case that becoming a mother in the private sector *causes* the woman to make career choices that increase her productivity. This could happen if the woman wishes to stay in her job and expects that becoming a mother may negatively affect her chances of continued employment and promotion in the future. To counter this she increases her productivity upon returning after giving birth. Another explanation could be compensation due to lack of job flexibility in the private sector: It is supposedly more costly for mothers than for nonmothers to hold stressful positions leading some mothers to select into the public sector as indicated by the mother indicator in the biprobit. Those staying in the private sector seem to be succesful in receiving compensation for the increase in disutility of work.

Two other potentially interesting parameters are the wage effects of sector change given motherhood. I.e.

$$E(Y_{10} - Y_{11}|C = 1, S = 1)$$

and

$$E(Y_{11} - Y_{10}|C = 1, S = 0).$$

Both sets of parameters are significant and similar in size as can be seen from table 8: Change of sector seems to be costly for mothers. In particular, mothers currently employed in the public sector would not seem to get a wage premium compared to non-mothers had they worked in the private sector since the decrease in wages caused by changing sector by far exceeds the small motherhood premium in the private sector.

[Table 8 around here]

Nielsen et al. (Forthcoming) also consider the effect of motherhood in the public and private sector and find results that seem to contradict the results from this analysis: They find that the effect of motherhood is positive in the public sector and negative in the private sector. The method differs from the one employed in this paper along many dimensions but the most important difference is that the estimated treatment effects are ATE. Hence the interpretation of the estimates from Nielsen et al. (Forthcoming) would be the effect of motherhood in a given sector for a woman drawn randomly from the population of women. Furthermore, when allowing for incomplete selection and conditioning on sector choice for a standardized woman they find that the most persistent log wage gap is seen in the counterfactual case; women who are actually public sector workers would, in particular, have been penalized for having children if they had been employed in the private sector. This corresponds well with the conclusion from this paper.

5 Discussion

We contribute to the existing literature on the effect of motherhood on wages by using high quality data and by implementing propensity score matching that in many respects is less restrictive than what has been used up until now. We discuss identification of direct and net effects of motherhood when covariates are likely to be affected by the choice of motherhood. In general, conditioning on covariates affected by the treatment is never justified, however under the assumption of linearity of the outcome equations this problem may be overcome. We take advantage of an exclusion restriction to perform regression adjusted matching. Furthermore, we extend the matching analysis to take account of discrete covariates that are affected by the treatment *without* assuming linearity of the outcome equation. By doing this we allow for heterogenous treatment effects across discrete endogenous covariates.

We estimate both net and direct effects of being a mother and find significantly negative impacts on wage growth for mothers. However, the direct effect is small. Hence, most of the difference can be explained by accounting for covariates that are likely to be affected by having children. We find that net effects of motherhood are twice as large in the public sector compared to the private sector. This may, however, be attributed to differences in choices between the two groups of mothers. The direct effect of motherhood is negative in the public sector and positive but small in the private sector. We argue that it is likely to be due to considerable job flexibility in the public sector.

Pinpointing the family gap is important if we are interested in equity and efficiency issues. Equity concerns both equal pay for equal work between mothers and non-mothers and between genders since differences in returns to having children contribute to explaining the gender wage gap. We would also suspect that for women to fully utilise their abilities - most definitely in the interest of society - they would have to be rewarded in accordance with their effort. To date, most empirical work has focused little on endogeneity issues when trying to identify direct effects. Furthermore, whether endogenous variables have any substantial impact on the parameter of interest is an empirical question. In our case allowing for sector specific treatment effects clearly changes the conclusion regarding costs of motherhood.

References

Albrecht, J. W., Edin, P., Sundström, M. & Vroman, S. B. (1999), 'Career Interruptions and Subsequent Earnings', *Journal of Human Resources* **34**, 294–311.

Angrist, J. D. & Evans, W. N. (1998), 'Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size', *American Economic Review* **88**, 450–477.

Black, D. A. & Smith, J. A. (Forthcoming), 'How Robust is the Evidence on the Effects of College Quality? Evidence from Matching', *Journal of Econometrics* .

Budig, M. J. & England, P. (2001), 'The Wage Penalty for Motherhood', *American Sociological Review* **66**, 204–225.

Butcher, K. & Case, A. (1994), 'The Effect of Sibling Composition on Women's Education and Earnings', *Quarterly Journal of Economics* **109**, 531–563.

Datta Gupta, N. & Smith, N. (2002), 'Children and Career Interruptions: The Family Gap in Denmark', *Economica* **69**(276), 609–629.

Dehejia, R. H. & Wahba, S. (1999), 'Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs', *Journal of the American Statistical Association* **94**(448), 1053–1062.

Ermisch, J. & Francesconi, M. (2001), 'Family Matters: The Impact of Family Background on Educational Attainments', *Economica* **68**, 137–156.

Heckman, J., Ichimura, H., Smith, J. & Todd, P. (1998), 'Characterizing Selection Bias Using Experimental Data', *Econometrica* **66**(5), 1017–1098.

Heckman, J., Ichimura, H. & Todd, P. (1997), 'Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme', *Review of Economic Studies* **64**, 605–654.

Heckman, J., Ichimura, H. & Todd, P. (1998), 'Matching as an Econometric Evaluation Estimator', *Review of Economic Studies* **65**, 261–294.

Heckman, J., LaLonde, R. & Smith, J. (1999), The Economics and Econometrics of Active Labor Market Programs, *in* A. Ashenfelter & D. Card, eds, 'Handbook of Labor Economics, Vol.', Vol. 3, Elsevier Science, Amsterdam, pp. 1865–2097.

Imbens, G. (2000), 'The Role of the Propensity Score in Estimating Dose Response Functions', *Biometrika* **87**(3), 706–710.

Jacobsen, J. P., Wishart Pearce III, J. & Rosenbloom, J. L. (1999), 'The Effects of Childbearing on Married Women's Labor Supply and Earnings: Using Twin Births as a Natural Experiment', *Journal of Human Resources* **34**, 449–474.

Jacobsen, J. & Rayak, W. (1996), 'Do Men Whose Wife Works Really Earn Less?', *American Economic Review* **86**(2), 268–273.

Klepinger, D., Lundberg, S. & Plotnick, R. (1999), 'How Does Adolescent Fertility Affect the Human Capital and Wages of Young Women?', *Journal of Human Resources* **34**, 421–448.

Korenman, S. & Neumark, D. (1991), 'Does Marriage Really Make Men More Productive?', *Journal of Human Resources* **26**(2), 282–307.

Korenman, S. & Neumark, D. (1992), 'Marriage, Motherhood, and Wages', *Journal of Human Resources* **27**, 233–255.

Lechner, M. (2001), *Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption*, Physica/Springer, chapter 3.

Loh, E. S. (1996), 'Productivity Differences and the Marriage Wage Premium for White Males', *Journal of Human Resources* **31**(3), 566–589.

Mincer, J. & Ofek, H. (1982), 'Interrupted Work Careers: Depreciation and Restoration of Human Capital', *Journal of Human Resources* **17**, 3–24.

Mincer, J. & Polachek, S. (1974), 'Family Investments in Human Capital: Earnings of Women', *Journal of Political Economy* pp. 76–108.

Nielsen, H. S., Simonsen, M. & Verner, M. (Forthcoming), 'Does the Gap in Family-Friendly Policies Drive the Family Gap?', *Scandinavian Journal of Economics* .

Phipps, S., Burton, P. & Lethbridge, L. (2001), 'In and Out of the Labour Market: Long-Term Income Consequences of Child-Related Interruptions to Women's Paid Work', *Canadian Journal of Economics* **34**, 411–429.

Robins, J. M. & Greenland, S. (1992), 'Identifiability and Exchangeability for Direct and Indirect Effects', *Epidemiology* **3**(2), 143–155.

Robinson, P. M. (1988), 'Root-N-Consistent Semiparametric Regression', *Econometrica* **56**, 931–954.

Rosenbaum, P. R. (1984), 'The Consequences of Adjustment for a Concomitant Variable That Has Been Affected by the Treatment', *Journal of the Royal Statistical Society Series A* **147**(5), 656–666.

Rosenbaum, P. R. & Rubin, D. B. (1985), 'Constructing a control group using multivariate matched sampling methods that incorporate the propensity score', *The American Statistician* **39**(1), 33–38.

Rosenbaum, P. & Rubin, D. (1983), 'The Central Role of the Propensity Score in Observational Studies for Causal Effects', *Biometrika* **70**(1), 41–55.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall/CRC.

Simonsen, M. & Skipper, L. (2003), Does High Tech Make a Difference? Estimating the Family Gap Using OLS, Matching, and Selection Models. Mimeo, University of Aarhus.

Smith, J. A. & Todd, P. (Forthcoming), 'Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?', *Journal of Econometrics* .

Waldfogel, J. (1998*a*), 'The Family Gap for Young Women in the United States and Britain: Can Maternity Leave Make a Difference?', *Journal of Labor Economics* **16**, 505–545.

Waldfogel, J. (1998*b*), 'Understanding the "Family Gap" in Pay for Women with Children', *Journal of Economic Perspectives* **12**, 137–156.

A Econometric Method

This appendix briefly explains the technical details and estimation procedures used in this paper. In order to be able to utilise (**A-3**) we need (**A-1**) and (**A-2**) to hold over the common support, $S_P = Supp\,(P\,|C=1) \bigcap Supp\,(P\,|C=0)$. See below for construction of this set.

Our matching estimator implemented takes on the following typical form

$$\Delta = \frac{1}{m_1} \sum_{i \in I_1 \cap S_P} \left(Y_{1i} - \sum_{j \in I_0} W(i,j) Y_{0j} \right), \tag{4}$$

where I_1 denotes the set of mothers, I_0 the set of women without children, and m_1 denotes the number of women in the set $I_1 \cap S_P$. Notice how the match for each mother $i \in I_1 \cap S_P$ is constructed as a weighted average over the outcomes of women without children, where the weights, $W(i,j)$, are constructed such that they depend on the distance between P_i and P_j. The different existing matching estimators differ in how the weights are constructed.

The standard estimators known as nearest neighbour matching use at most one person from the comparison group in constructing a match for the mothers. **Kernel matching** and **Local Linear Regression matching** are nonparametric matching estimators that construct matches for each mother using kernel weighted averages over multiple women without children. Relative to the simple nearest neighbor matching, by using kernel techniques we reduce the variance of our matching estimates by making use of information from additional women. As a cost, we introduce small sample bias because of the increased distance between mothers and matched women, measured in terms of their probabilities of having a child. The weights used in the kernel matching estimates are

$$W(i,j) = \frac{K\left(\frac{P_i - P_j}{h_n}\right)}{\sum_{k \in I_0} K\left(\frac{P_i - P_k}{h_n}\right)} = \frac{K_{ij}}{\sum_{k \in I_0} K_{ik}},$$

where K is a kernel function and h_n is a bandwidth. Similarly, the weight for local linear matching is

$$W(i,j) = \frac{K_{ij} \sum_{k \in I_0} K_{ik} \left(P_k - P_i\right)^2 - \left[K_{ij}\left(P_j - P_i\right)\right] \left[\sum_{k \in I_0} K_{ik}\left(P_k - P_i\right)\right]}{\sum_{j \in I_0} K_{ij} \sum_{k \in I_0} K_{ik}\left(P_k - P_i\right)^2 - \left(\sum_{k \in I_0} K_{ik}\left(P_k - P_i\right)\right)^2} \tag{5}$$

We use this latter local linear matching in the analyses above instead of the simpler kernel matching because of its desirable statistical properties (see Heckman et al. (1997) for details). Specifically, one of the advantages of the local linear estimator over simple kernel weighting is that the local linear estimator will converge at a faster rate at boundary points, and, as was depicted in figure 1, both groups have values of P very close to the boundary value at zero, and there are values of P at the boundary point at one for mothers. Hence, we will be able to put more confidence in the local linear version of matching compared to kernel matching. Due to support considerations (see Heckman, Ichimura, Smith & Todd (1998) for practical details) we use as kernel function, K, the biweight kernel given by

$$K(s) = \begin{cases} \frac{15}{16}(1-s^2)^2 & \text{for } |s| < 1 \\ 0 & \text{otherwise} \end{cases}$$

and a bandwidth following Silverman (1986) "rule of thumb".

To estimate $\boldsymbol{\beta}_0$ in (2) we apply partially linear regression as in Heckman et al. (1997): Firstly, we estimate the propensity to have children, $P(\mathbf{Z})$ by standard probit including the exclusion restriction. We then estimate β_0 by

$$\begin{aligned} \hat{\beta}_0 \quad = \quad & \left(\sum_{\{i:d_i=0\}} \left(\mathbf{x}_i - \hat{E}[\mathbf{x}_i|\hat{P}(\mathbf{z}_i)]\right) \left(\mathbf{x}_i - \hat{E}[\mathbf{x}_i|\hat{P}(\mathbf{z}_i)]\right)' \hat{I}_i \right)^{-1} \cdot \\ & \sum_{\{i:d_i=0\}} \left(\mathbf{x}_i - \hat{E}[\mathbf{x}_i|\hat{P}(\mathbf{z}_i)]\right) \left(y_i - \hat{E}[y_i|\hat{P}(\mathbf{z}_i)]\right)' \hat{I}_i \end{aligned}$$

where $\hat{P}(\mathbf{Z})$ is used to nonparametrically estimate $E[Y_0|P, C = 0]$ and $E[\mathbf{X}|P, C = 0]$ using local linear regression techniques. $\hat{I}$ is an indicator function, $\mathbf{1}_{[\hat{f}(\hat{P}|D=0)>\hat{c}_{0q}]}$, that trims away a small fraction of observations ($q\%$) to ensure uniform consistency of our estimators. The estimated density of the propensity is obtained using the biweight kernel and $\hat{c}_{0q}$ satisfies

$$\sup_{\hat{c}_{0q}} \frac{1}{J} \sum_{\{i:d_i=0\}} 1[\hat{f}(\hat{P}|C=0) < \hat{c}_{0q}] \leq q$$

where J is the cardinality of $\{i : c_i = 0\}$. We then remove $\mathbf{X}\widehat{\boldsymbol{\beta}}_0$ from both Y_1 and Y_0 and use the resulting residuals in place of the corresponding $Y's$ in (1) using (5) as weights.

B Balancing Score

Our use of the propensity score in reducing the curse of dimensionality should balance the distribution of covariates, $\mathbf{X}$, see Rosenbaum & Rubin (1983). This property is investigated informally in the present secion. We do not investigate the full distribution of covariates, but focus merely on the first two moments (see also Dehejia & Wahba (1999)). Table A presents the characteristics of the matching estimators implemented. The standardized differences are calculated as (see Rosenbaum & Rubin (1985))

$$100 * \frac{\left(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_{0M}\right)}{\sqrt{\left(S_1^2 + S_{0M}^2\right)/2}},$$

where $\overline{\mathbf{X}}_1$ denotes the sample mean of mothers, $\overline{\mathbf{X}}_{0M}$ denotes the sample mean of the matched non-mothers, and $S_1^2 + S_{0M}^2$ is the sum of squared bootstraped standard errors.

The table below demonstrates the balancing properties. The balancing performance of the estimator seems to be very high: In no case do the standardized differences in means for covariates exceed 7%. Thus, matching on the the propensity score including only limited cross terms does seem to achieve balancing of covariates.

C Sensitivity Analysis

We choose the optimal bandwith using cross validation as in Black & Smith (Forthcoming). Due to out large sample our estimates are not sensitive to the choice of bandwith or kernel and thus using Silverman rule of thumb to choose the bandwidth (Silverman (1986)) instead of cross validation does not matter. Also, performing nearest neighbour matching both with and without replacement give results that are exactly identical to the local linear matching analysis. This is remarkable given that the estimates are based on very different subsamples but obviously because of the large estimation sample.

It is possible that the treatment effect varies with age. One possible test is to calculate ATET conditional on completed fertility. We perform the matching analysis for the part of the sample aged 39-40. Women aged 39 or older give birth to 2.7% of all children born in Denmark in 1997. Some of these children are obviously not the women's first child. Hence, choosing 39 years of age as the cut-off point does not seem too unreasonable. We do not find significant changes in the estimated family gaps.

To check for internal consistency of our model, we estimate the average treatment effect on the untreated (TUT): The effect of having children in terms of wages for the group of women who do not have children. We find that our estimates of TUT do not differ from the estimates of ATET. I.e. it is not the case that women who do not have children would benefit in terms of wages for having so indicating a serious problem with the procedure for accounting for unobserved characteristics. All results are available from the author's on request.

Notes

[1]The family gap may exist for several reasons: mothers may invest differently in household production. This can result in career interruptions in the form of leaves or in permanent withdrawal from the labour market. They may also have different preferences for working conditions such as unplanned overtime. Differences in bargaining power may also cause wage differences between mothers and non-mothers: Mothers are more likely to be geographically tied if moving children is perceived as costly. Hence the value of mothers' outside option is reduced. Employers are expected to know this and mothers' bargaining power is reduced accordingly. Finally, discrimination may explain a potential wage gap.

[2]Note, though, that the estimated parameter may not be representative of the effect of having an extra child in general.

[3]Potential experience is defined in the literature as (age-years of schooling-age at school start)

[4]Most often, children from separated homes live with their mother. Moreover, in 1996 the number of adoptions amounted to only 600 in total. In our sample this would amount to approximately 30 adoptions in 1996.

[5]Notice that these assumptions are generally weaker than the usual "strong ignorability" assumptions of Rosenbaum & Rubin (1983), which are often both invoked in the literature of matching even though they are overly strong given that the parameter of interest be the treatment on the treated.

[6]See appendix A for ways of constructing this set.

[7]Note that strictly speaking we do not need the probit to be the exact propensity score. Only do we need it to balance the distributions of our attributes - hence we need it to be a balancing score. For details, see appendix B.

[8]Note that this variable is our exclusion restriction. It is not included when estimating the propensity used to perform non-residual based matching. The remaining coefficients are not sensitive to the exclusion of the information on number of siblings.

[9]We define number of wrong predictions as $\sum_{i=1}^{n}\left(D_i - \mathbf{1}_{(\widehat{p}_i \geq .5)}\right)^2$.

[10]Remember that Efrons R^2 is defined as

$$1 - \frac{\sum_{i=1}^{n} (D_i - \widehat{p}_i)^2}{\sum_{i=1}^{n} \left(D_i - \frac{\sum_{j=1}^{n} D_j}{n} \right)^2}$$

[11]See Simonsen & Skipper (2003) for an in depth discussion of the gains from propensity score matching in this particular set up.

[12]See Appendix C for a full list of the conditioning variables used in regression adjusted matching along with coefficients from estimating the partial linear regression.

[13]We reestimate the propensities for each sector separately. The coefficients are similar to those in Table 2 and are available from the authors on request.

[14]To evaluate the predictive power of the bivariate probit we calculate the number of right predictions by comparing the predicted state, characterised by the maximum of the predicted probabilites over the 4 states, to the real state. In 54.4% of the times the model predicts the real state. This should be compared to an expected 25% had we had no model.

TABLE 1

SELECTED MOMENTS, DESCRIPTIVE STATISTICS

Variables[b]	All	Mothers	Non-mothers
Log wages	4.80	4.81	4.79
	(0.28)	(0.26)	(0.30)
Age (years)	30.71	33.87	26.88
	(5.80)	(4.20)	(5.13)
Experience (years)	7.27	9.39	4.70
	(4.87)	(4.29)	(4.25)
Length of completed education (years)	12.23	12.20	12.26
	(2.45)	(2.42)	(2.48)
Type of highest completed education:			
General (0/1)	0.22	0.19	0.26
Business (0/1)	0.34	0.33	0.35
Industry (0/1)	0.01	0.01	0.01
Construction (0/1)	0.01	0.01	0.01
Graphical (0/1)	0.01	0.01	0.01
Services (0/1)	0.02	0.02	0.02
Food and beverages (0/1)	0.04	0.04	0.03
Agricultural (0/1)	0.01	0.01	0.01
Transportation (0/1)	0.00	0.00	0.00
Teaching (0/1)	0.06	0.09	0.04
Humanities (0/1)	0.03	0.03	0.03
Musical (0/1)	0.00	0.00	0.00
Social (0/1)	0.04	0.03	0.05
Technical (0/1)	0.02	0.01	0.02
Public security (0/1)	0.00	0.00	0.00
Unknown (0/1)	0.08	0.07	0.09
Owner of real estate (0/1)	0.42	0.55	0.26
Married (0/1)	0.43	0.66	0.15
Province (0/1)	0.63	0.68	0.57
Private sector (0/1)	0.52	0.46	0.60
High level occupation (0/1)	0.10	0.10	0.10
Medium level occupation (0/1)	0.21	0.23	0.18
Low level occupation (0/1)	0.52	0.51	0.53
Total duration of unemployment (weeks)	64.20	87.40	36.20
	(94.40)	(106.90)	(66.60)
Number of siblings when 15-17 years of age	1.16	1.00	1.35
	(1.02)	(1.05)	(0.95)
Number of siblings missing (0/1)	0.19	0.23	0.14
Sample size	29210	16012	13198

[a] Standard deviations shown in parenteses.

[b] Omitted educational type is 'health care', omitted occupation category is lowest level.

TABLE 2

Coefficient Estimates and Asy. Std. Err. from Motherhood Probit

Dep. variable: 1 for Mothers, 0 for Non-Mothers

Full Sample, 16,012 Mothers and 13,198 Non-Mothers

Variables[a]	Coeff	Asy Std Error
Intercept	**−12.59**	0.40
Age (years)	**0.76**	0.02
Age squared	**−0.01**	0.00
General (0/1)	0.37	0.24
Business (0/1)	**−1.83**	0.23
Industry (0/1)	1.19	1.00
Construction (0/1)	0.36	1.16
Graphical (0/1)	−0.56	1.41
Services (0/1)	0.18	0.59
Food and beverages (0/1)	−0.47	0.41
Agricultural (0/1)	−0.53	0.46
Transportation (0/1)	−3.26	2.69
Teaching (0/1)	0.27	0.50
Humanities (0/1)	**−1.22**	0.51
Musical (0/1)	−0.94	1.19
Social (0/1)	**−2.78**	0.52
Technical (0/1)	−0.91	0.63
Public security (0/1)	**−0.70**	1.56
Unknown (0/1)	**−0.97**	0.24
Length of education	**−0.11**	0.01
General (0/1)*length	**−0.08**	0.02
Business (0/1)*length	**0.13**	0.02
Industry (0/1)*length	−0.13	0.08
Construction (0/1)*length	−0.05	0.10
Graphical (0/1)*length	0.11	0.12
Services (0/1)*length	−0.04	0.05
Food and beverages (0/1)*length	0.02	0.03
Agricultural (0/1)*length	0.02	0.03
Transportation (0/1)*length	0.18	0.21
Teaching (0/1)*length	−0.03	0.03
Humanities (0/1)*length	0.05	0.03
Musical (0/1)*length	0.04	0.07
Social (0/1)*length	**0.14**	0.03
Technical (0/1)*length	−0.03	0.04
Public security (0/1)*length	0.45	0.25
Unknown (0/1)*length	−0.03	0.02
Province	**0.27**	0.02
No. siblings	**0.07**	0.01
No. siblings missing	**0.24**	0.03

[a]The omitted type of education is health care.

TABLE 3

COMPARISON OF ESTIMATED CROSS-SECTIONAL ATET

Dep. Variable: Log Hourly Wage Rate in 1997

Full Sample, 16,012 Mothers and 13,198 Non-Mothers

	Differences in Means	Net Effect	Net Effect	Direct Effect
		Local Linear Matching[a,b]	Local Linear Regression Adjusted Matching[a,c] Potentially end. excl.	Local Linear Regression Adjusted Matching[a,d] Potentially end. incl.
$E\left(Y_1 - Y_0 \vert C=1\right)$	**0.016**	**−0.066**	**−0.070**	**−0.015**
	(0.004)	(0.005)	(0.005)	(0.004)

[a] Densities were estimated using a biweight kernel and a bandwidth chosen via cross validation.
The overlapping support region was determined using a 2 % trimming rule and a biweight kernel function. Std. errors are based on 99 bootstraps with 100% resampling. See Heckman, Ichimura & Todd (1997) for details.

[b] 16,003 mothers and 12,059 non-mothers used

[c] Regression adjustment includes educational categories and length. 16,003 mothers and 12,039 non-mothers used.

[d] Regression adjustment includes experience, educational categories and length, occupational categories, and interruptions from the labour market. 16,004 mothers and 12,039 non-mothers used.

TABLE 4

NUMBER OF MOTHERS AND NON-MOTHERS IN EACH SECTOR

	Mothers, $C = 1$	Non-Mothers, $C = 0$
Public Sector, $S = 0$	8602 (62,0%)	5280 (38,0%)
Private Sector, $S = 1$	7410 (48,3%)	7918 (51,7%)

TABLE 5
COMPARISON OF ESTIMATED CROSS-SECTIONAL ATET
SECTOR CHOICE EXOGENOUS
Dep. Variable: Log Hourly Wage Rate in 1997
Full Sample, 16,012 Mothers and 13,198 Non-Mothers

	Differences in Means	Net Effect	Direct Effect
		Local Linear Matching[a,b]	Local Linear Regression Adjusted Matching[a,c] Potentially end. incl.
$E(Y_{10} - Y_{00} \mid C = 1, S = 0)$	-0.003	$\mathbf{-0.065}$	$\mathbf{-0.035}$
	(0.005)	(0.005)	(0.006)
$E(Y_{11} - Y_{01} \mid C = 1, S = 1)$	**0.040**	$\mathbf{-0.053}$	0.006
	(0.005)	(0.006)	(0.006)

[a]Densities were estimated using a biweight kernel and a bandwidth chosen via cross validation. The overlapping support region was determined using a 2 % trimming rule and a biweight kernel function. Std. errors are based on 99 bootstraps with 100% resampling. See Heckman et al. (1997) for details.
[b]8,594 mothers and 4,739 non-mothers used in public sector calculations, 7,140 mothers and 7,322 non-mothers used in private sector calculation.
[c]Regression adjustment includes experience, educational categories and length, occupational categories, and interruptions from the labour market. 8,594 mothers and 4,739 non- mothers used in public sector, 7,410.mothers and 7,313 non-mothers used in private sector calculations.

TABLE 6

COEFFICIENT ESTIMATES AND ASY. STD. ERR.
MOTHERHOOD AND SECTOR BIVARIATE PROBIT

Dep. variables: 1 for Mothers, 0 for Non-Mothers, 1 for Private Sector, 0 for Public Sector
Full Sample, 16,012 Mothers, 13,198 Non-Mothers, 15,328 in Private, and 13,882 in Public Sector

-

	Motherhood		Sector Choice	
Variables[a]	Coeff.	Asy. Std. Error	Coef.	Asy. Std. Error
Intercept	**−12.51**	0.38	**−1.17**	0.20
Motherhood			**−0.44**	0.03
Age (years)	**0.76**	0.02		
Age squared	**−0.01**	0.00		
General (0/1)	0.34	0.23	**3.05**	0.22
Business (0/1)	**−1.84**	0.23	**2.35**	0.22
Industry (0/1)	1.29	1.08	**2.54**	0.90
Construction (0/1)	0.39	1.24	1.42	1.06
Graphical (0/1)	−1.54	1.31	**2.89**	1.31
Services (0/1)	0.19	0.57	0.55	0.51
Food and beverages (0/1)	−0.48	0.40	**3.88**	0.40
Agricultural (0/1)	−0.52	0.46	**3.01**	0.45
Transportation (0/1)	−3.24	2.11	0.83	2.43
Teaching (0/1)	0.27	0.51	**−2.13**	0.92
Humanities (0/1)	**−1.22**	0.54	7.56	0.56
Musical (0/1)	−0.92	1.22	**3.79**	1.15
Social (0/1)	**−2.73**	0.55	**5.34**	0.54
Technical (0/1)	−0.90	0.63	**3.17**	0.62
Public security (0/1)	−7.67	4.78	**0.92**	4.78
Unknown (0/1)	**−0.98**	0.23	**2.39**	0.23
Length of education	**−0.11**	0.01	**0.03**	0.01
General (0/1)*length	**−0.07**	0.02	**−0.17**	0.02
Business (0/1)*length	**0.13**	0.02	**−0.07**	0.02
Industry (0/1)*length	−0.14	0.09	−0.06	0.08
Construction (0/1)*length	−0.05	0.10	0.03	0.09
Graphical (0/1)*length	0.10	0.11	−0.09	0.11
Services (0/1)*length	−0.05	0.05	**0.10**	0.04
Food and beverages (0/1)*length	0.02	0.03	**−0.22**	0.03
Agricultural (0/1)*length	0.02	0.03	**−0.12**	0.03
Transportation (0/1)*length	0.17	0.17	0.09	0.20
Teaching (0/1)*length	−0.03	0.03	0.09	0.06
Humanities (0/1)*length	0.05	0.03	**−0.38**	0.03
Musical (0/1)*length	0.04	0.08	**−0.19**	0.07
Social (0/1)*length	**0.14**	0.03	**−0.25**	0.03
Technical (0/1)*length	−0.03	0.04	**−0.10**	0.04
Public security (0/1)*length	0.50	0.34		
Unknown (0/1)*length	−0.03	0.02	**−0.09**	0.02
Province	**0.27**	0.02	−0.02	0.02

TABLE 6, CONTINUED

Coefficient Estimates and Asy. Std. Err.
Motherhood and Sector Bivariate Probit

Dep. variables: 1 for Mothers, 0 for Non-Mothers, 1 for Private Sector, 0 for Public Sector
Full Sample, 16,012 Mothers, 13,198 Non-Mothers, 15,328 in Private, and 13,882 in Public Sector

	Motherhood		Sector Choice	
Variables[a]	Coef.	Asy. Std. Error	Coef.	Asy. Std. Error
No. siblings	**0.07**	0.01		
No. siblings missing	**0.25**	0.03		
Father empl. in public sector			**−0.07**	0.03
Mother empl. in public sector			**−0.11**	0.02

	Coef.	Asy. Std. Error
Correlation between error terms, ρ	**0.15**	0.02

[a]The omitted type of education is health care.

TABLE 7

Comparison of Estimated Cross-Sectional ATET
Sector Choice endogenous
Dep. Variable: Log Hourly Wage Rate in 1997
Full Sample, 16,012 Mothers and 13,198 Non-Mothers

	Net Effect	Direct Effect
	Local Linear Matching[a,b]	Local Linear Regression Adjusted Matching[a,c] Potentially end. incl.
$E\left(Y_{10}-Y_{00} \mid C=1, S=0\right)$	**−0.110**	**−0.044**
	(0.008)	(0.007)
$E\left(Y_{11}-Y_{01} \mid C=1, S=1\right)$	**−0.050**	**0.018**
	(0.006)	(0.006)

[a] Densities were estimated using a biweight kernel and a bandwidth chosen via cross validation. The overlapping support region was determined using a 2 % trimming rule and a biweight kernel function. Std. errors are based on 99 bootstraps with 100% resampling. See Heckman et al. (1997) for details.

[b] 8,058 mothers and 5,270 non-mothers used in public sector calculations, 7,390 mothers and 7,361 non-mothers used in private sector calculation.

[c] Regression adjustment includes experience, educational categories and length, occupational categories, and interruptions from the labour market. 8,058 mothers and 5,270 non- mothers used in public sector, 7,393.mothers and 7,331 non-mothers used in private sector calculations.

TABLE 8

COMPARISON OF ESTIMATED CROSS-SECTIONAL ATET
SECTOR CHOICE ENDOGENOUS
Dep. Variable: Log Hourly Wage Rate in 1997
Full Sample, 16,012 Mothers and 13,198 Non-Mothers

	Net effect	Direct Effect
	Local Linear Matching[a,b]	Local Linear Regression Adjusted Matching[a,c] Potentially end. incl.
$E\left(Y_{11}-Y_{10}\|C=1,S=0\right)$	**−0.091**	**−0.085**
	(0.005)	(0.004)
$E\left(Y_{10}-Y_{11}\|C=1,S=1\right)$	**−0.084**	**−0.084**
	(0.004)	(0.004)

[a]Densities were estimated using a biweight kernel and a bandwidth chosen via cross validation. The overlapping support region was determined using a 2 % trimming rule and a biweight kernel function. Std. errors are based on 99 bootstraps with 100% resampling. See Heckman et al. (1997) for details.

[b]7,361 mothers and 7,993 non-mothers used in public sector calculations, 7,367 mothers and 8,051 non-mothers used in private sector calculation.

[c]Regression adjustment includes experience, educational categories and length, occupational categories, and interruptions from the labour market. 7,374 mothers and 7,999 non- mothers used in public sector, 7,360.mothers and 8,063 non-mothers used in private sector calculations.

TABLE A1

Covariate Imbalance in Local Linear Matching Sample

Means and Standardized Differences in Percentage Points[a]

Variables	Means, Mothers	Matched Means, Non-Mothers	% Bias
Age (years)	33.87	33.87	0.21
Age squared	1165.07	1164.67	0.15
General (0/1)	0.19	0.21	−4.41
Business (0/1)	0.33	0.31	3.98
Industry (0/1)	0.01	0.01	−0.25
Construction (0/1)	0.01	0.01	0.42
Graphical (0/1)	0.01	0.01	−1.03
Services (0/1)	0.02	0.02	−0.38
Food and beverages (0/1)	0.04	0.04	−0.85
Agricultural (0/1)	0.01	0.01	−2.48
Transportation (0/1)	0.09	0.08	−1.38
Teaching (0/1)	0.09	0.08	2.38
Humanities (0/1)	0.03	0.03	0.37
Musical (0/1)	0.00	0.00	0.19
Social (0/1)	0.03	0.03	0.56
Technical (0/1)	0.01	0.01	−0.87
Public security (0/1)	0.00	0.00	0.06
Unknown (0/1)	0.07	0.08	−4.35
Length of education	12.20	12.06	6.02
General (0/1)*length	1.94	2.11	−4.07
Business (0/1)*length	3.74	3.52	4.07
Industry (0/1)*length	0.11	0.11	−0.12
Construction (0/1)*length	0.12	0.11	0.53
Graphical (0/1)*length	0.11	0.12	−1.04
Services (0/1)*length	0.24	0.24	−0.14
Food and beverages (0/1)*length	0.52	0.53	−0.68
Agricultural (0/1)*length	0.12	0.15	−2.20
Transportation (0/1)*length	0.01	0.02	−1.31
Pedagoghic (0/1)*length	1.28	1.18	2.48
Humanistic (0/1)*length	0.43	0.41	0.42
Musical (0/1)*length	0.01	0.01	0.12
Social (0/1)*length	0.05	0.44	0.55
Technical (0/1)*length	0.21	0.23	−0.91
Public security (0/1)*length	0.02	0.02	0.14
Unknown (0/1)*length	0.71	0.81	−3.70
Province	0.68	0.68	−0.45

[a]Sample size is 16,003 mothers and 12,059 non-mothers.

TABLE A2

Coefficients from Partial Linear Regression

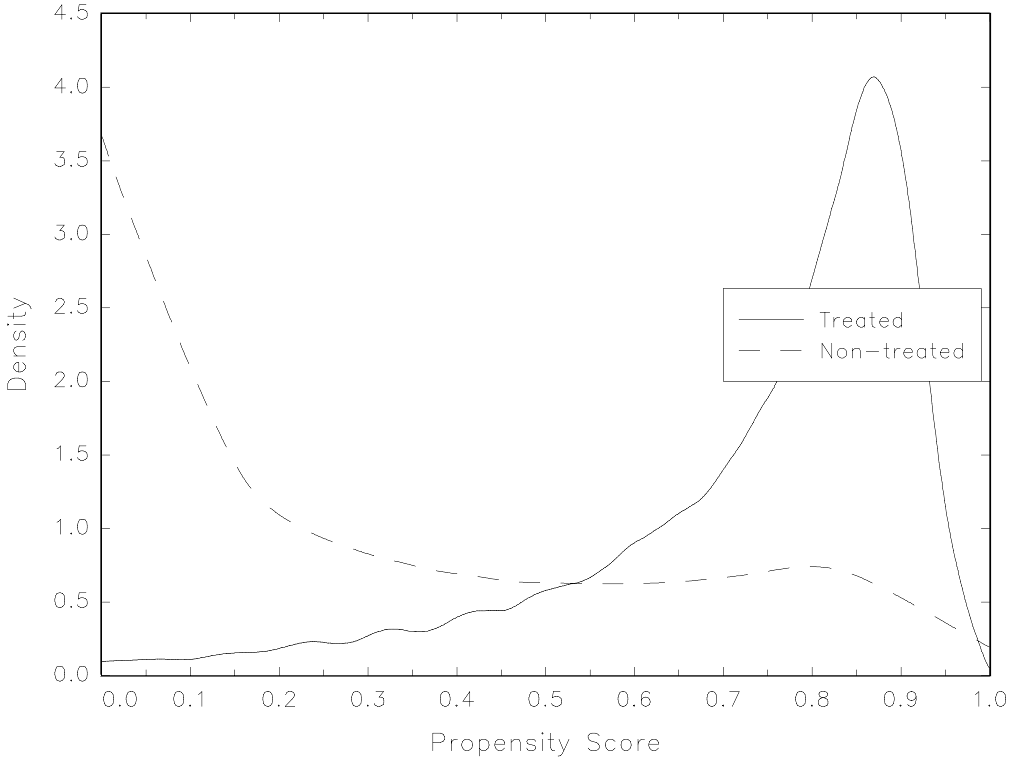
Variables	Coefficients	Std. deviations
Experience (years)	−0.008	0.005
Experience squared	0.001	0.000
General (0/1)	0.320	0.079
Business (0/1)	0.537	0.089
Industry (0/1)	0.485	0.360
Construction (0/1)	0.874	0.490
Graphical (0/1)	0.426	0.408
Services (0/1)	0.318	0.221
Food and beverages (0/1)	0.310	0.099
Agricultural (0/1)	0.380	0.093
Transportation (0/1)	0.821	1.680
Pedagoghic (0/1)	0.691	0.312
Humanistic(0/1)	0.345	0.258
Musical (0/1)	1.789	1.081
Social (0/1)	0.710	0.218
Technical (0/1)	0.281	0.384
Public security (0/1)	−0.854	1.482
Unknown (0/1)	0.361	0.081
Length of education	0.052	0.005
General (0/1)*length	−0.019	0.005
Business (0/1)*length	−0.044	0.008
Industry (0/1)*length	−0.037	0.030
Construction (0/1)*length	−0.073	0.042
Graphical (0/1)*length	−0.023	0.037
Services (0/1)*length	−0.026	0.017
Food and beverages (0/1)*length	−0.024	0.010
Agricultural (0/1)*length	−0.027	0.008
Transportation (0/1)*length	−0.061	0.139
Teaching (0/1)*length	−0.050	0.019
Humanities (0/1)*length	−0.026	0.015
Musical (0/1)*length	−0.109	0.060
Social (0/1)*length	−0.043	0.013
Technical (0/1)*length	−0.017	0.026
Public security (0/1)*length	0.060	0.100
Unknown (0/1)*length	−0.027	0.006
Province	−0.093	0.014

TABLE A2 CONTINUED

Coefficients from Partial Linear Regression

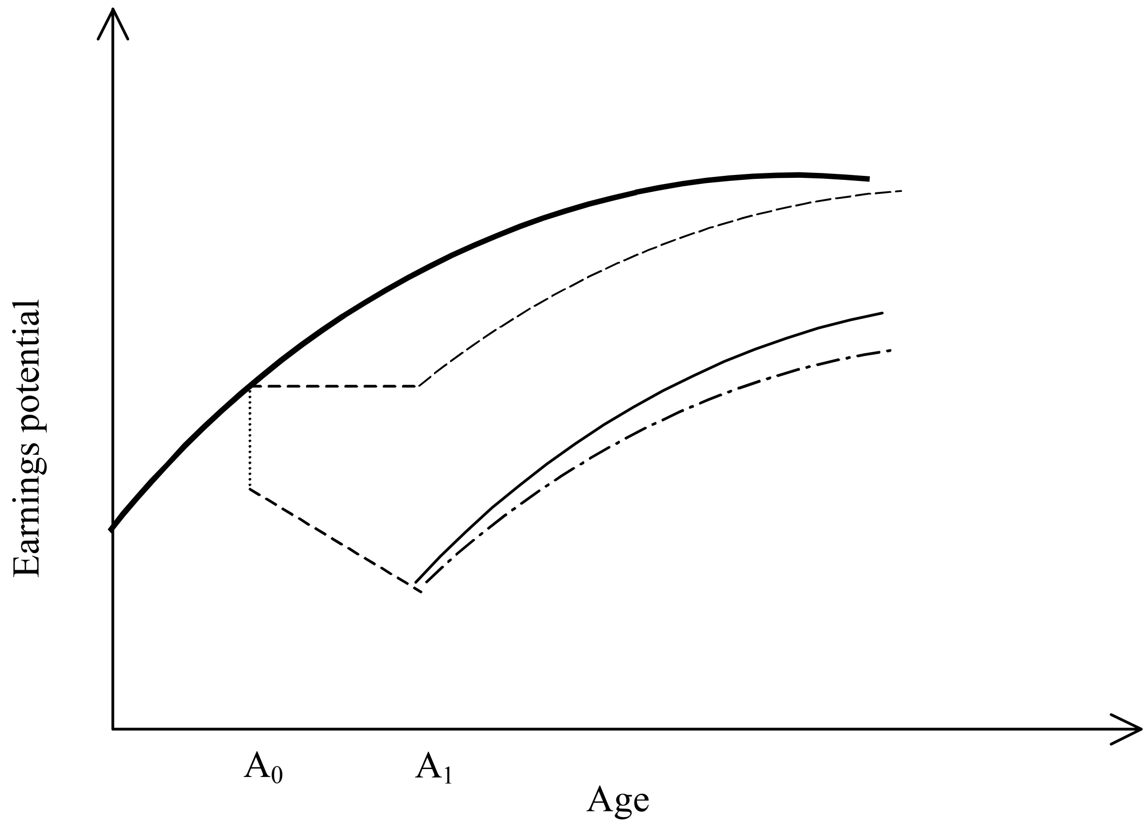
Variables	Coefficients	Std. deviations
Private Sector (0/1)	0.052	0.016
Owner	0.023	0.007
Married	−0.015	0.012
High Occupation	0.116	0.023
Medium Occupation	−0.042	0.027
Low Occupation	−0.051	0.021
Latest unempl. spell 1981 (0/1)	−0.024	0.069
Latest unempl. spell 1982 (0/1)	0.034	0.096
Latest unempl. spell 1983 (0/1)	0.031	0.043
Latest unempl. spell 1984 (0/1)	0.007	0.040
Latest unempl. spell 1985 (0/1)	−0.007	0.044
Latest unempl. spell 1986 (0/1)	0.011	0.039
Latest unempl. spell 1987 (0/1)	0.030	0.024
Latest unempl. spell 1988 (0/1)	−0.005	0.042
Latest unempl. spell 1989 (0/1)	0.001	0.026
Latest unempl. spell 1990 (0/1)	0.004	0.029
Latest unempl. spell 1991 (0/1)	−0.014	0.026
Latest unempl. spell 1992 (0/1)	−0.013	0.017
Latest unempl. spell 1993 (0/1)	−0.022	0.020
Latest unempl. spell 1994 (0/1)	−0.021	0.019
Latest unempl. spell 1995 (0/1)	−0.029	0.018
Latest unempl. spell 1996 (0/1)	−0.037	0.010
Latest non-part. spell 1981 (0/1)	0.020	0.038
Latest non-part. spell 1982 (0/1)	0.004	0.032
Latest non-part. spell 1983 (0/1)	0.014	0.053
Latest non-part. spell 1984 (0/1)	0.024	0.034
Latest non-part. spell 1985 (0/1)	0.008	0.022
Latest non-part. spell 1986 (0/1)	0.003	0.032
Latest non-part. spell 1987 (0/1)	0.022	0.030
Latest non-part. spell 1988 (0/1)	0.015	0.043
Latest non-part. spell 1989 (0/1)	−0.007	0.015
Latest non-part. spell 1990 (0/1)	0.001	0.009
Latest non-part. spell 1991 (0/1)	−0.002	0.022
Latest non-part. spell 1992 (0/1)	−0.036	0.025
Latest non-part. spell 1993 (0/1)	−0.043	0.013
Latest non-part. spell 1994 (0/1)	−0.056	0.023
Latest non-part. spell 1995 (0/1)	−0.093	0.023
Latest non-part. spell 1996 (0/1)	−0.074	0.025
Total duration of unempl. spells (years)	−0.001	0.002
Total duration of non-part. spells (years)	−0.015	0.006

FIGURE 1



Nonparametric estimates of propensity densities (biweight kernel), probit $P(\mathbf{Z})$ model; bandwidth using Silverman (1986) rule of thumb. —— Mothers, - - - Non-mothers

FIGURE 2



Earnings Potential. —— No interruptions, no family gap; - - - Experience foregone, no family gap; —— Experience foregone, family gap, depreciation, catching up; · — ·— Experience foregone, family gap, depreciation, no catching up.