Gender Differences in Job Satisfaction and Labour Market Participation: UK Evidence from Propensity Score Estimates

Anna Sanz de Galdeano*
European University Institute

Preliminary version

Abstract

In contrast with the extensive evidence on women's disadvantaged position in the labour market, British female employees report higher levels of job satisfaction than their male counterparts. The goal of this paper is to find out whether this differential reflects a true causal relationship between gender and job satisfaction or if it is simply a correlation due to systematic differences in personal and job characteristics by gender and/or to a sample selection problem. To this purpose, the effect of interest is estimated applying parametric estimation techniques as well as propensity score methods. It is found that there is a significant positive effect of being female on job satisfaction, which is robust to the different estimation methods used. A proposed explanation relies on expectations. If job outcomes are evaluated relative to expectations, women's higher job satisfaction may reflect their lower expectations regarding their jobs, which temselves likely result from the poorer position in the labour market that women have traditionally held.

JEL Classification: C14; J16; J22; J28.

Keywords: Job satisfaction; Gender; Propensity score methods; Matching.

1 Introduction

This paper explores the relationship between job satisfaction and gender. Previous empirical research into job satisfaction has shown that British women consistently express themselves as more satisfied with their jobs than men. This is a surprising finding, given

*Present address: Department of Economics, European University Institute, Via dei Roccettini 9, I-50016 San Domenico di Fiesole (FI), Italy. E-mail: galdeano@iue.it

the disadvantaged position and more restricted role that women have traditionally held in the labour market.

The goal of this emprical study is to find out whether there is a true causal effect of gender on job satisfaction. If gender were randomly assigned to the population of workers, the observed gender satisfaction differential could be interprted as a causal effect. However, this is obviously not the case and neither controlled randomization is feasible nor there exists a convincing natural experiment. In particular, complicating the analysis is the existence of two sources of biases. First, there are substantial differences in terms of personal and job characteristics by gender. If characteristics such as age, education or hours of work vary with gender, this may help explain the observed gender satisfaction differential.

The second reason for biases is that there may as well be a problem of sample selection: as fewer women than men work, it may be the case that dissatisfied men are working, while potentially dissatisfied women occupy some other labour force status. Then, women's higher job satisfaction would have a statistical (a sample selection problem) rather than causal interpretation.

In order to account for these biases, I estimate the effect of interest using a variety of statistical methods, which allows me to avoid reliance on a single method of inference and to asses robustness of the results. First, I use parametric estimation methods including ordinary least squares, ordered probability models and the Heckman selection model. Second, propensity score methods are applied, illustrating how to calculate the average effect of treatment in the population by stratifying and matching on the basis of the estimated propensity score. Propensity score methods offer a way to estimate treatment effects when controlled randomization is impossible and there are no convincing natural experiments providing a substitute to randomization. The estimated propensity score (defined as the probability of assignment to tratment, conditional on covariates) is a scalar variable on the unit interval that summarizes the observable variables, offering a diagnostic on the comparability of the treatment and comparison obsevations and quickly revealing the extent of overlap in the two groups. The assumption underlying this methodology, called the unconfoundedness assumption, is that assignment to treatment is associated only with observable variables. In other words, the relevant differences between female and male employees are captured in the observable covariates. Although this is a strong assumption and its plausibility clearly depends on the richness of the data set used, the empirical evidence in favour of these methods is compelling.

It is found that there is a significant positive effect of being female on job satisfaction which is robust to the different estimation methods applied. Furthermore, this effect is not reduced even when its estimation is based on the most careful matching of male and female employees (that is, pairing together female and male employees that are almost identical in terms of their observable characteristics). Based on this statistical evidence, I interpret the results as estimates of the causal effect of gender on job satisfaction.

The rationalisation and interpretation of the existence of a positive causal relationship between being female and job satisfaction is outside the scope of this paper. Clark's (1997) proposed explanation appeals to the notion of relative well-being, in particular relative to employees' expectations. If job outcomes are evaluated relative to expectations, women's higher job satisfaction may reflect their lower expectations regarding their jobs, which temselves likely result from the poorer position in the labour market that they have traditionally held. Evidence supporting this interpretation has been found by Clark (1997) and Sanz de Galdeano (2001).

The article is organised as follows. Section 2 presents a review of the relevant literature on job satisfaction and gender up to now. Section 3 sets up the basic staitistical framework. The data set used is described in Section 4. Sections 5 and 6 describe the alternative estimation tehìchniques used (parametic estimation and propensity score methods, respectively) as well as their advantages and drawbaks and present the empirical results obtained. A sensitivity analysis is developed in Section 6. Section 7 concludes the article.

2 Literature Review

Job satisfaction, while the subject of popular attention and of an extensive sociological and industrial psychology literature, has been studied by relatively few economists. Freeman (1978) suggested that this probably "reflects the professional suspicion of what is a subjective variable (a variable that measures what people say rather than what people do) and, indeed, one that purports to measure individual utility" (p. 135).

From an economic perspective, the main problem lies in satisfaction data being comparable across individuals: one person does not use the same scale of answers to job satisfaction questions in the same way as another. Thus, the limitations of such data exist and should not be ignored.[1] However, there are three main arguments that support the credibility and robustness of analysis drawn from subjective assessments on job satisfaction.

Firstly, there is a huge literature within psychology that takes seriously survey responses on feelings of well-being. As Clark and Oswald (1996) and Blanchflower and Oswald (1997) remark, psychologists are likely to be more skilled than economists in judging the quality

[1] Bertrand and Mullainathan (2001) summarise the experimental and empirical literature that has investigated the meaningfulness of answers to subjective questions and integrate the main findings into a measurement error framework.

of such data and it would be extreme to argue that economists know more about human psychology than do psychologists.

Secondly, as Clark (1997) points out, "it is not clear why whole groups, such as women or older workers, should systematically understand the satisfaction scale so differently" (p. 344).

Last, there are strong correlations, in the expected directions, between job satisfaction and poor mental health (Wall et al., 1978), length of life (Palmore, 1969) and coronary heart disease (Sales and House, 1971). Furthermore, there is a well documented relationship between job satisfaction and labour market behaviour. Clark et al. (1998), Akerlof et al. (1988), McEvoy and Cascio (1985) and Freeman (1978) all find that job satisfaction predicts future quits, while Clegg (1983) and Mangione and Quinn (1975) show that job satisfaction responses are negatively correlated with absenteeism and positively correlated with workers' productivity. If the answers people give to job satisfaction questions were purely idiosyncratic, it is hard to see how the above results could have been obtained.

Therefore, the existence of such correlations suggests that the analysis of workers' subjective well-being provides a number of insights into certain labour market behaviours, which may explain the noticeable rise in the number of articles examining job satisfaction in recent years.

The economic literature on job satisfaction to date can be divided into studies considering the workforce as a whole,[2] those which compare the job satisfaction of trade union members relative to non-members,[3] those which focus on race[4], age[5] or gender,[6] one that focuses on establishment size[7] and one that examines the evolution of the distribution of job satisfaction over time.[8]

This paper focuses on the relationship between job satisfaction and gender. Previous empirical research into job satisfaction in the labour market has shown that women consistently express themselves as more satisfied with their jobs than men. Women's higher job satisfaction has been reported in recent work by Clark (1996, 1997), Sloane and Williams (2000), and Blanchflower and Oswald (1992). This is a surprising finding, given that there is extensive evidence that women hold a clearly disadvantaged position in the labour market

[2]See for example, Clark and Oswald (1996), Clark (1996), Freeman (1978) and Hamermesh (1977).

[3]See for example, Sloane and Bender (1998), Meng (1990), Miller (1990), Berger at al. (1983), Freeman (1984) and Borjas (1979).

[4]See for example, Bartel (1991).

[5]See for instance, Clark, Oswald and Warr (1996).

[6]See for example, Clark (1997), Sloane and Williams (2000) and Sloane and Ward (2000).

[7]Idson (1990).

[8]Hamermesh (2001).

in terms of earnings[9] and in areas such as hiring/firing[10] and promotion.[11] Furthermore, other research has found that women report higher levels of stress in their life.[12]

A number of potential explanations have been put forward to try to explain why the responses to survey measures of job satisfaction differ by gender. To begin with, a variety of personal and job characteristics are found to vary with gender and to have significant effects on the levels of job satisfaction. If a gender effect is present in characteristics such as age, education, hours of work, establishment size and occupation, this may help to explain a gender effect in job satisfaction. Clark (1997) uses data from the first wave (year 1991) of the British Household Panel Survey and finds that this array of controls rarely explains the higher reports of job satisfaction by women.

Secondly, it has been suggested that men and women may have different values concerning which factors are important in their job and this may explain the gender satisfaction gap. Clark (1997) finds that male workers within the British Household Panel Survey rank promotion prospects, pay and job security more highly than women, but women are more likely to mention good relations with managers, the actual work itself and the hours of work. Furthermore, an emphasised importance of pay is associated with lower reports of job satisfaction and emphasis on relations at work with higher reports of job satisfaction. Work values do not, however, help to explain the gender effect in job satisfaction, which remains basically unchanged once the work values variables have been included in the regressions.

A third reason suggested to explain the gender job satisfaction gap relies on sample selectivity. In the present context, the argument is that if dissatisfied female workers find it easier to leave the market place than equally dissatisfied male workers, the remaining females will have higher average job satisfaction, since the sample is now biased. If this were the case, then women's higher job satisfaction would have a statistical (a sample selection problem) rather than causal interpreation. The technique employed by Clark (1997) to deal with this problem is a full maximum likelihood estimation of the Heckman selection model. When the sample selection term was included in the job satisfaction equations it turned out to be insignificant and had little effect on the coefficients of the independent variables. Therefore, Clark (1997) concludes that sample selection does not account for women's higher job satisfaction in the data set he uses.

Finally, Clark (1997) rationalises the existence of a positive relationship between being

[9] See for example, Blau and Kahn (1992) and Wright and Ermish (1991).

[10] See for example, Riach and Rich (1997).

[11] See for example, Lazear and Rosen (1990) and Weiler (1990).

[12] See for example, Clark and Oswald (1994).

female and job satisfaction as a reference level effect.[13] The basis for this interpretation, taken from the social psychology literature, is the finding that individuals tend to evaluate experiences relative to some kind of norm or reference level, which has been interpreted as an individual's expectations about his/her job. Hence, women's higher job satisfaction may reflect their lower expectations regarding their jobs, which themselves likely result from the poorer position in the labour market that they have traditionally held. Both Clark (1997) and Sanz de Galdeano (2001) provide some evidence supporting this interpretation coming from cross-sectional and panel analyses respectively.

3 Statistical Framework

The goal of this study is to answer the following questions: is there a causal effect of gender on job satisfaction?. Or the observed gender differential in job satisfaction is simply a correlation due to systematic differences in individual and job characteristics between men and women and/or to a sample selection problem?

All efforts to infer causal effects must confront the fundamental problem of causal inference: one wants to compare outcomes across males and females, but one can only observe each individual being either female or male (Holland, 1986). Furthermore, labour economists often have to deal with the *selection problem*: "*this is the problem of identifying conditional probability distributions from random sample data in which the realizations of the conditioning variables are always observed but the realizations of outcomes are censored.*" (Manski, 1995, p.22)

When studying the determinants of job satisfaction, available surveys such as the British Household Panel Survey, provide background data for each respondent and job satisfaction data for those respondents who work. Even if all subjects respond fully to the questions posed, there remains a censoring problem in that the surveys do not provide job satisfaction data for respondents who do not work. When comparing job satisfaction levels, almost random samples can be drawn from the male population if almost all men work. This is often not possible for females, since not all women work and job satisfaction data on non working women are not available.

Let each member of the population be characterized by the vector (Y, F, W, X, D), where Y is the outcome to be predicted (job satisfaction), F is a female dummy variable that

[13] There is a wide variety of reference level effects, and the ways and degrees to which they can influence behaviour is the object of an extensive social psychology literature. Reference levels have variously been formulated as the outcome or payoff that could have been received had uncertainty been resolved differently (regret theory: Bell, 1982), other individuals' outcomes, actions or behaviours (equity theory: Adams, 1963 and social exchange theory: Homans, 1961) prior expectations (disappointment theory: Bell, 1985), etc.

takes the value one if the individual is female and the value zero otherwise, W represents the individual's wage and X denotes other conditioning variables. The variable D takes the value one if Y is observed (that is, if the individual is employed) and the value zero otherwise. Let a random sample be drawn from the population. One observes all realizations of (F, W, X, D), but observes Y only when $D = 1$. In the absence of further assumptions, censoring makes it impossible to learn anything about the expected value $E(Y|F, W, X)$ of Y conditional on F, W, X. To see this, write $E(Y|F, W, X)$ as the sum

$$E(Y|F, W, X) = E(Y|F, W, X, D = 1)P(D = 1|F, W, X) +$$
$$E(Y|F, W, X, D = 0)P(D = 0|F, W, X)$$

Now the selection problem becomes apparent as only $E(Y|F, W, X, D = 1)$, $P(D = 1|F, W, X)$, and $P(D = 0|F, W, X)$ can be identified, while $E(Y|F, W, X, D = 0)$ can by definition never be observed. As a consequence, without further assumptions, standard methods are unable to identify $E(Y|F, W, X)$ because necessary observations are inherently missing.

To cope with this selection problem early suggestions were to assume exogenous or ignorable selection. That is,

$$E(Y|F, W, X, D = 1) = E(Y|F, W, X, D = 0)$$

This assumption implies that the censored outcomes have the same distribution as the observed ones, conditional on F, W and X. In particular, it implies that $E(Y|F, W, X)$ coincides with the observable distribution $E(Y|F, W, X, D = 1)$. However, the credibility of this assumption diminished sharply when economists began a sustained effort to understand the role of self-selection in determining when behavioural outcomes are observed. In particular, the work of labour economists studying market wage determination has been very influential. Labour economists reasoned as follows (see Gronau, 1974):

1. Wage data are available only for respondents who work.

2. Respondents work when they choose to do so.

3. The wage one would be paid influences the decision to work.

4. Therefore the distribution of observed and unobserved wages may differ.

Following the standard reservation wage model, it is conventionally assumed that each individual knows the wage W that would be paid if he or she were to work. The individual chooses to work if W is greater than some lowest acceptable wage R, called the person's reservation wage, and chooses not to work if W is below the reservation wage. So job satisfaction levels are observed when $W > R$ and are censored when $W < R$.[14] Let $D^* = W - R$ be a latent variable that represents the unobservable criterion followed by individual concerning his/her participation into employment. The observed variable D is defined as:

$$\begin{aligned} D &= 1 \text{ if } D^* > 0 \\ D &= 0 \text{ otherwise} \end{aligned}$$

Assume that the following population relationship, henceforth called structural equation, between the explanatory variables F, W, X and the outocome variable Y exists:

$$Y = \alpha F + \beta W + \gamma X + \varepsilon \tag{1}$$

where ε is a random error.

To complete the model, define $W - R = \delta Z + u$ and let the selection equation be expressed as

$$D = \mathbf{1}(\delta Z + u > 0) \ \in \{0, 1\} \tag{2}$$

with Z selection regressors, that may or may not be different from F, W and X, and a random error term u.

The probability of being an employee is likely correlated with the wage one would be paid and with the individual's potential job satisfaction. As fewer women than men work, it may be the case that dissatisfied men are working, while potentially dissatisfied women choose to occupy some other labour force status. The hypothesis that women's reservation wage is higher relies on the observation[15] that women, in effect, have more alternative uses for their time (market work, home work and leisure) than do men, who for the most part divide their time between two uses, market work and leisure. Hence, women have non market outside options that are on average better than those for men and, therefore, their opportunity cost of working is higher.[16]

The job satisfaction regression function for the subsample of available data is

$$E(Y|F, W, X, D = 1) = \alpha F + \beta W + \gamma X + E(\varepsilon|u > -\delta Z) \tag{3}$$

[14] The reservation wage model does not predict whether a person works if $W = R$, but it is conventionally assumed that this event occurs with probability zero in the population.

[15] See, for instance, Mincer (1962, 1963).

[16] As Killingsworth and Heckman (1986) point out, this argument does not explain why home work is primarily women's work, but, in the context of this framework, it does at least suggest why women's valuation of time spent out of employment is higher.

In the case of independence between ε and u, selection is ignorable or exogenous, the conditional mean $E(\varepsilon|u > -\delta Z)$ is 0 and the regression function for the selected subsample is the same as the population regression function. Least squares estimators may be used to estimate on the selected subsample and the only cost of having an incomplete sample is a loss of efficiency.

In the general case, however, the conditional mean $E(\varepsilon|u > -\delta Z)$ is nonzero. Hence, regression estimators of the parameters of equation 1 fit on the selected sample omit the final term of equation 3 as a regressor, so that the selection bias takes the form of an omitted variable specification error. Intuitively, since women have on average better outside market options than men, it may be that potentially dissatisfied women were less likely to be in employment than equally dissatisfied men. If this were the case, the observed gender satisfaction differential would have a statistical explanation (a sample selection problem) rather than reflecting a true causal relationship.

4 The Data

The data used in this paper are taken from waves 1-8 of the British Household Panel Survey[17] (BHPS hereafter), covering the period 1991-1998. The BHPS is an annual survey of each adult member of a nationally representative sample of around 5,000 households across Great Britain, giving a total of approximately 10,000 interviews. The same individuals are re-interviewed at each wave. The main topics of the BHPS include household organisation, labour market participation, education and training, opinion-type questions exploring the socio-economic values of respondents, etc. The survey consists of a short household-level questionnaire and a more detailed individual questionnaire. All the variables used in this study come from the individual questionnaire.

As previously pointed out, job satisfaction data are only available for those respondents who work. The employment ratey for the pooled sample of employees in the period 1991-98 is 69.98%, clearly higher for men at 75.45% than for women at 64.98%.

The basic sample for the analyses presented below has been selected from the original sample on the basis of three criteria. First, individuals aged 18-65 are kept. This filter leaves a sample size of 63,478 individual-year observations, which is considered as the starting sample. Second, the self-employed are excluded due to the inapplicability of some of the questions used in the study, such as the ones referring to establishment size or to the number of hours normally worked per week. This filter reduces the sample size by 9.3%, leaving 57,537 individual-year observations. The third criterion requires the exclusion of

[17] For extensive information on the BHPS, see http://www.iser.essex.ac.uk/bhps.

all the observations for which one or more of the variables used in the analysis is missing. As a result, the sample size is further reduced to 33,857 individual-year observations, which implies an additional 41.1% loss. Neither male nor female respondents appear to be significantly more susceptible to loss of observations in this selection step.

The final (unbalanced) panel consists of 16,934 and 16,923 individual-year observations on 4,333 and 4,406 male and female respondents respectively.[18] A remarkable feature of this panel, which time structure by gender is described in Table 1, is that it contains a few more female than male respondents, despite the fact that the female employment rate is lower. This is due to the fact that, when excluding the self-employed from the analysis in the second selection step, many more male than female workers were dropped.

4.1 Job Satisfaction Data

The job satisfaction questions in the BHPS asked workers to give an integer response on a scale of 1 to 7 that best described how satisfied or dissatisfied they were with seven[19] specific facets of their present job, with 1 representing the lowest level of job satisfaction and 7 the highest. Finally, individuals were asked "All things considered, how satisfied or dissatisfied are you with your present job overall using the same 1-7 scale?". The job satisfaction measures used in the empirical analyses are the overall one and two of the specific ones: job satisfaction with total pay and with the actual work itself.

Table 2 displays the mean level of overall job satisfaction for the entire sample of employees as well as the percentage of employees who ranked their satisfaction levels according to each of the possible sale responses. It is interesting to note that the percentage of employees who declare themselves as satisfied at work (reporting satisfaction of 5, 6 or 7 on the 1-7 scale) is remarkably high (80.16%). Thus, British employees seem well satisfied with their jobs overall.[20].

In addition, the reported levels of job satisfaction by female and male employees and t-statistics for the null hypotheses that the averages for the two groups are identical are also presented. The mean level of the satisfaction score for women is 5.56 as against only 5.21 for men. This difference is statistically significant with a p-value smaller than 0.00001. Moreover, 21.69% of women reported satisfaction of 7 on the 1-7 scale, as against only 13.39% of men. Therefore, the message to be learnt from Table 2, consistent with the

[18] These numbers correspond to overall job satisfaction. The number of observations varies slightly with the satisfaction measure used.

[19] However, three of these job satisfaction measures (satisfaction with promotion prospects, use of initiative and relations with boss) are not available for 1998.

[20] This finding is in line with that of Clark (1996).

findings of previous studies on job satisfaction and gender, is clear: British women report, on average, higher levels satisfaction at work.

4.2 Explanatory Variables

The Data Appendix contains detailed definitions of all the individual and job related variables included as controls in the regressions,[21] whereas Table 3 presents summary statistics for all of them. For each variable the mean and standard deviation for the entire sample of employees are given. In addition, the mean and standard deviation for male and female employees as well as t-statistics from the tests of identical means by gender are reported.

Overall there are substantial differences between the two groups in terms of most of the explanatory variables. In particular, there are striking differences in terms of important covariates such as education, income, hours of work and occupational category among others. The existence of such differences highlights the need for the careful statistical adjustment procedures described in Section 6.

In the next two sections, I describe the statistical methods carried out on this data set as well as their advantages and drawbacks and I discuss the results obtained.

5 Parametric Estimation Methods

5.1 Single Equation Models

As a starting point and as a benchmark for later comparisons, equation 1 is estimated by using OLS as well as an ordered probit model, which takes account of the ordinal nature of the outcome variable[22] [Table 4, columns (1) and (2) respectively]. The following individual and job related characteristics are included as controls in the regressions: age, age-squared/1000, education dummies, health dummies, marital status dummies, establishment size dummies, occupational category dummies, log income, log hours, second job, full time job, temporary worker, union member, region dummies and industry dummies.

The estimated coefficients on the female dummy variable are 0.227 (OLS) and 0.179 (ordered probit model[23]), and they are both statistically significant with a p-value smaller

[21] Work values variables have not been used in the analyses because they are only available for 1991. However, it has been checked that the results obtained do not change substantially when restricting the analyses to 1991 and including work values variables in the regressions.

[22] See for instance, Zavoina and McKelvey (1975) and Greene (1993). For a brief outline of the various alternative methods for analysing job satisfaction measured in ordinal scales see Sloane and Williams (1994).

[23] It is important to note that the coefficients of an ordered probability model, like those of any non linear regression model, are not necessarily the marginal effects one is accostumed to analysing. It is obvious that these will vary with the values of the covariates. However, coefficient and marginal effects have the

than 0.0001. Hence, according to these estimates, there is a positive effect of being female on job satisfaction. The next subsection investigates whether this effect remains once sample selection is accounted for by using a Heckman selection model.

5.2 Heckman Selection Model

The traditional sample selection model in econometrics began with the work of Heckman (1974) on wages and labour supply and was developed, expanded and elaborated further in a series of papers in the late 1970s by Heckman (1978, 1979) and others.

Consider the sample selection model composed by the structural and selection equations presented in the previous section (equations 1 and 2 respectively). It is further assumed that ε and u are distributed bivariate normal with means zero, variances equal to σ and 1 respectively, and with correlation ρ.

The job satisfaction regression function for the subsample of available data is

$$E(Y|F,W,X,D=1) = \alpha F + \beta W + \gamma X + E(\varepsilon|u > -\delta Z) = \alpha F + \beta W + \gamma X + \theta\lambda(\delta Z) \tag{4}$$

where $\theta = \sigma\rho$ and $\lambda(\delta Z) = f(\delta Z)/F(\delta Z)$ is the inverse of Mill's ratio, and where f and F are the unit normal p.d.f. and c.d.f. respectively. Given the result in 4, consistent estimates of the parameters of ineterest can be obtained either estimating 1 and 2 by maximum likelihood or by a two-step procedure in which, in the first stage, probit estimates of 2 are used to compute the inverse of Mill's ratio which are used in the second stage to estimate 1 by least squares.

Full maximum likelihood and two-step estimates of the Heckman selection model are presented in Table 4. The control variables included in the structural equation are the same as those entering the previous OLS and ordered probit regressions. Similarly to Clark (1997), I estimate the selection equation using a probit model for being an employee that includes all of the non-job control variables previously used plus a number of other variables that are assumed not to influence reported levels of job satisfaction. They include partner's pay, partner's hours of work, the provision of care for others and the number of own children in the household, all of them interacted with gender. Unlike Clark (1997), and following Hamermesh (1999), main effects are present for every variable included in interaction terms.

The coefficients on the female dummy variable estimated by ML and the two-step procedure are 0.208 and 0.195. They are somewhat smaller than the OLS benchmark

same sign, and positive coefficients in an ordered probability model are associated with a higher estimated probability that the individual is satisfied at work.

estimate of 0.227, but they remain positive and statistically significant with a p-value smaller than 0.0001.

On the basis of these estimates, one would be tempted to conclude that sample selection does not account for women's higher job satisfaction and interpret the estimates as evidence confirming the existence of a positive effect of being female on job satisfaction. However, some caveats should be borne in mind when interpreting these results. Heckman's procedure was used in the mid-1970s with widespread enthusiasm and still today it is probably the most widely used approach to correcting for selection bias. Nevertheless, empirical practice in labour economics has seen a decline in the use of this method, the reason of which is twofold. First, estimates tend to be unstable, non-robust and sensitive to minor changes in the specification of the X and Z vectors. Second, the distributional assumptions of bivariate normality may be false and, specially if Z does not differ from F, W and X (that is, if there is no exogenous variable which determines the selection into employment and, at the same time, is excluded from the job satisfaction equation[24]), identification of the model is made on the basis of an arbitrary distributional assumption.

The results obtained have been checked for robustness by the use of different specifications of the employment probit. However, it is hard to think of a variable that could realistically determine the probability of employment without affecting job satisfaction. In fact, it seems reasonable to believe that characteristics such as partner's pay and hours of work or the number of own children in the household may well have an influence on a subjective variable like job satisfaction. Furthermore, there are strong theoretical reasons to belive that these variables and labour supply might be jointly determined.[25]

In light of these criticisms and in order to assess robustness of the results, next section introduces and implements an alternative semi-parametric methodology based on the propensity score.

6 Propensity Score Methods

The propensity score methodology was originally developed by Rosenbaun and Rubin (1983) and it was recently applied by Dehejia and Wahba (1998, 1999) to the data originally used by Lalonde (1985). Other recent applications in economics include Heckman, Ichimura, Smith and Todd (1997), Imbens, Rubin and Sacerdote (1999), and Persson, Tabellini and Trebbi (2001). Propensity score methods offer a diagnostic on the compara-

[24]This excluded component is sometimes referred to as an *instrumental variable*.

[25]Angrist and Evans' (1998) paper, for instance, is motivated by the possibility that fertility and labour supply may be jointly determined.

bility of the treatment and comparison obsevations and provide a way to estimate treatment effects when controlled randomization is impossible and there are no convincing natural experiments providing a substitute to randomization, as it is the case in this application. The theory behind this estimation strategy is briefly summarised in the next subsection.

6.1 An Introduction to the Propensity Score Methodology

Using the potential outcome notation, (Rubin, 1974), let $Y_i(1)$ represent the value of the outcome (job satisfaction) when unit i is exposed to regime 1 (called treatment), and let $Y_i(0)$ represent the value of the outcome when unit i is exposed to regime 0 (called control). There is a fundamental missing data problem, since only one of $Y_i(1)$ or $Y_i(0)$ can be observed for any unit, because one cannot observe the same unit under both treatment and control. Let T_i be a binary indicator for the treatment status (1 if unit i is subject to treatment, 0 if it is exposed to the control). Then, the realized outcome for unit i is $Y_i = T_i \cdot Y_i(1) + (1 - T_i) \cdot Y_i(0)$. In order to maintain the same terminology as in the evaluation literature, in what follows female employees are defined as 'treated' while male employees make up the 'control' group.[26] The treatment effect for unit i is $\omega_i = Y_i(1) - Y_i(0)$.

The aim of this study is to find out whether there is a causal effect of gender on job satisfaction. In the context of the potential outcomes framework, the counterfactual question to be answered is the following: would the level of job satisfaction of an employee be different if he/she had the opposite gender? Therefore, the effect of interest is the population average treatment effect, that is, $ATE = E\left[Y_i(1) - Y_i(0)\right]$, where the first term is only observed for the treated and the second one is only obsereved for the control units. This missing data problem would be solved if gender were randomly assigned to the populatiom of workers,[27] but this is obviously not the case. Finally, let X_i be a vector of observable covariates. This vector includes a wide range of individual and job related characteristics such as age, education, health status, marital status, establishment size, weekly hours of work, etc.[28] The key assumption underlying the propensity score methodology is the unconfoundedness assumption.

Definition 1 *UNCONFOUNDEDNESS*

[26] Logically, one could also define the male employees as 'treated' and the female employees as 'controls'. In this case, the estimates obtained would have the absolute value, but its sign and therefore its interpretation would be the opposite.

[27] Randomization implies that $T_i \perp Y_i(1), Y_i(0)$. Thus, $E[Y_i(1)] = E[Y_i|T_i = 1]$ and $E[Y_i(0)] = E[Y_i|T_i = 0]$.

[28] Detailed definitions of all the variables used in the statistical analyses are given in the Data Appendix.

Assignment to treatment T_i is unconfounded, given pre-treatment variables X_i, if

$$T_i \perp Y_i(1), Y_i(0) \mid X_i$$

This is a strong assumption, since it asserts that assignment to treatment is associated only with observable variables. In other words, the relevant differences between treated and control units are captured in the observable covariates and, conditional on them, assignement to treatment can be taken to be random. Although this assumption is debatable and its plausibility clearly depends on the richness of the data set used, the empirical evidence in favour of the propensity score methodology is compelling and encouraging. In fact, when Dehejia and Wahba (1999) apply it to Lalonde's[29] (1986) data set for a range of propensity score specifications and estimators, they obtain estimates of the treatment effect that are much closer to the experimental estimate than Lalonde's nonexperimental estimates.

The unconfoundedness assumption validates comparisons for units with the same value of the covariates:

$$\begin{aligned} E[Y_i(1)|T_i &= 1, X_i] = E[Y_i(1)|T_i = 0, X_i] = E[Y_i(1)|X_i] \\ E[Y_i(0)|T_i &= 0, X_i] = E[Y_i(0)|T_i = 1, X_i] = E[Y_i(0)|X_i] \end{aligned}$$

Hence, the fundamental missing data problem that is always confronted when trying to infer causal effects in non experimental settings disappears and the population average treatment effect is then identified,

$$ATE = E\left[Y_i(1) - Y_i(0)\right] = E\left[E[Y_i(1)|T_i = 1, X_i] - E[Y_i(0)|T_i = 0, X_i]\right],$$

where the outer expectation is taken over the distribution of X_i.

In principle, one could stratify data into cells, each defined by a particular value of X_i, and compute the ATE by averaging the differences in outcomes of treated and comparison units of the same X_i characteristics.[30] However, in practice it can be difficult to estimate $E\left[Y_i(1) - Y_i(0)\right]$ in this manner when the covariates, X_i, are high dimensional.[31] This motivated the work by Rosenbaum and Rubin (1983), who proposed an alternative based

[29] Lalonde (1986) estimated the impact of a labour training program on postintervention income levels. He used data from a randomized evaluation of the program and examined the extent to which nonexperimental estimators can replicate the unbiased experimental estimate of the treatment impact when applied to a composite data set of experimental treatment units and nonexperimental comparison units.

[30] The regression equivalent of this procedure would require the inclusion of a full set of non parametric interactions between all the observable covariates.

[31] For example, if all n variables are dichotomous, the number of possible values for the vector X will be 2^n. As the number of variables increases, the number of cells will increase exponentially, augmenting the probability of finding cells containing only female or only male individuals.

on the propensity score that circumvents the necessity to condition on the entire set of covariates.

Definition 2 *PROPENSITY SCORE, ROSENBAUM AND RUBIN (1983)*

The propensity score is the conditional probability of receiving the treatment given the pre-treatment variables:

$$p(X_i) \equiv \Pr(T_i = 1|X_i)$$

The propensity score has two important properties:

Lemma 1 *BALANCING OF PRE-TREATMENT VARIABLES GIVEN THE PROPENSITY SCORE, ROSENBAUM AND RUBIN (1983)*

$$T_i \perp X_i \mid p(X_i)$$

Proof. See Statistical Appendix ■

Combined with unconfoundedness, the balancing property leads to the key property of the propensity score:

Lemma 2 *UNCONFOUNDEDNESS GIVEN THE PROPENSITY SCORE, ROSENBAUM AND RUBIN (1983)*

Suppose that assignment to treatment is unconfounded. Then assignment to treatment is unconfounded given the propensity score:

$$T_i \perp Y_i(1), Y_i(0) \mid p(X_i)$$

Proof. See Statistical Appendix ■

This result implies that instead of having to condition on the entire set of covariates, it is sufficient to condition on a scalar variable, the propensity score. Therefore, it follows immediately from the unconfoundedness assumption and from Lemma 2 that:

$$ATE = E\left[Y_i(1) - Y_i(0)\right] = E\left[E[Y_i(1)|T_i = 1, p(X_i)] - E[Y_i(0)|T_i = 0, p(X_i)]\right], \quad (5)$$

where the outer expectation is taken over the distribution of $p(X_i)$.

The next subsection describes the steps for the implementation of the propensity score methodology.

6.2 Estimation of the Propensity Score

The first step in the implementation of this methodology is to estimate the propensity score, which of course is not known. I estimate it by following a simple algorithm proposed by Dehejia and Wahba (1998), which is discussed in further detail in the Statistical Appendix. A probit model is used, but the results are not sensitive to the choice of other standard models. Lemma 1 (the balancing property) forms the basis of the algorithm used to estimate the propensity score. In other words, observations are grouped into strata defined on the estimated propensity score and it is checked whether the covariates are balanced across the treated and control observations within each stratum. Interaction and higher order terms are added and blocks are divided into finer blocks until this balance is achieved.

Graph 1 presents the histogram of the estimated propensity score for BHPS treated and control units. Logically, the number of treated (control) observations increases (decreases) as the propensity score increases (decreases). The first bin, for instance, contains only 41 treated units, whereas the last bins contains only 177 control units. However, there is always a minimal overlap between treated and control observations, which allows one to proceed with estimation.

The next step consists in estimating the average effect of treatment in the population given the estimated propensity score. It follows from 5 that an unbiased estimate of the treatment effect arises from conditioning on $p(X_i)$, which entails exact matching on $p(X_i)$. This is unfeasible in practice, since it is extremely rare to find two units with exactly the same propensity score. There are, however, several alternative and feasible procedures based on stratifying and matching (nearest and radius methods of matching) on the basis of the estimated propensity score.

6.3 The Stratification Estimator

The stratification estimator relies on the same strata defined when estimating the propensity score. Therefore, in each stratum the covariates are balanced between treated and control units by construction. The effect of gender on job satisfaction is estimated by summing the within stratum difference in means between the treated and control observations (that is, female and male employees), where the sum is weighted by the number of observations within each stratum [Table 5, row (1), column (1)]. Exact expressions for the stratification estimator and its standard errors are given in the Statistical Appendix. An alternative is a within block regression of job satisfaction on covariates, again taking a weighted sum over the strata [Table 5, columns (2) and (3)], which can help eliminate the remaining within-block differences (although such a regression should not have an important effect when the covariates are well balanced). The stratification estimates from a difference in means and linear regression adjustment are 0.261 and 0.176, not far from the benchmark OLS and Heckman ML estimates of 0.226 and 0.208 respectively.

6.4 The Matching Estimators

In the matching method, each treated (control) unit is matched with replacement to the control (treated) unit such that their propensity scores are sufficiently close to be considered as being approximately the same. Matching with replacement means that many treated (control) units can be matched to the same control (treated) unit.

Consider first the nearest-match method, which consists in matching each treated (control) unit to the control (treated) unit(s) with the closest propensity score. No unit is discarded with this method, since all units have a nearest match, no matter how far it is. However, it may be the case that, for some units, the nearest-match is too far to consider the conditioning on $p(X_i)$ in 5 approximately valid. Therefore, to the extent that the motivation for matching is to condition on $p(X_i)$, one may be willing to admit less units in the analyses in order to get more accurate matches in exchange. The radius method of matching consists in matching each treated (control) unit to the control (treated) units whose propensity scores are within some tolerance level δ. If a treated (control) unit has no control (treated) units within a δ-radius, this unit is discarded. Hence, in switching from the nearest-match to the radius method one improves the quality of the matches but ends up using less observations.

The effect of interest is estimated averaging the differences in job satisfaction between the female and male matched employees. [Table 5, column (1)]. As in the stratification estimation, an alternative is to perform a weighted regression of job satisfaction on covariates [Table 5, rows (3) to (6), columns (2) and (3)]. Exact expressions for the matching estimators, its standard errors and the associated weights are given in the Statistical Appendix. The first thing to note is that for both the nearest match and the radius method of matching, controlling for covariates has very little impact on the estimates, confirming that the covartiates are well balanced across the matched treated and control units. The regression adjusted estimates range from 0.284 (nearest match) to 0.243 (for a radius of $\delta = 0.00001$), slightly decreasing as the radius is reduced. The smaller the chosen radius is, the bigger the standard errors are, facing a potential bias-variance tradeoff. However, for all the values of δ, the radius method of matching yields statistically significant estimates in the application.

In conclusion, in this application propensity score matching and stratification methods yield very similar results to those from the parametric estimation techniques previously applied and confirm the existence of a positive and significant effect of being female in job satisfaction. Based on this collage of evidence, this effect is interpreted causally.

7 Sensitivity Analysis

7.1 Sensitivity to the Specification of the Propensity Score

The specification previously chosen for the propensity score (that is, the simplest one satisfying Lemma 1) is linear. It is important to check that the results do not change substantially when using the propensity score in a nonlinear functional form. Several sensitivity checks[32] in which higher order terms and interaction variables were added have demonstrated that the estimates of the effect of interest are indeed robust to the specification used for the propensity score.

7.2 Other Measures of Job Satisfaction

All the statistical work previously presented has used overall job satisfaction as a summary measure of satisfaction with all aspects of work. It is, however, interesting to check whether the results obtained are paticularly sensitive to estimation using other alternative measures of job satisfaction, such as satisfaction with total pay and with the actual work itself, which, as pointed out by Clark (1997), capture 'satisfaction with an extrinsic aspect of the job and with an intrinsic reward' (p. 353) respectively.

Tables 6 and 7 replicate Table 5 using job satisfaction with total pay and with the actual work itself as dependent variables. The resulting coefficients when using satisfaction with pay are somewhat bigger than those obtained when using the overall sastisfaction measure. On the other hand, the coefficients obtained when using satisfaction with the work itself are slightly smaller. However, in the case of both satisfaction measures the estimates remain always positive and statistically significant with a p-value smaller than 0.0001.

8 Conclusion

In this paper I report estimates of the effect of gender on job satisfaction. In order to account for potential biases and to avoid reliance on a single method of inference I estimate the effect of interest applying a variety of statistical methods. First, I use parametric estimation methods including ordinary least squares, ordered probability models and Heckman

[32]The results of these analyses, not reported in the paper, are available upon request from the author.

selection model. Second, propensity score methods are applied, illustrating how to calculate the average effect of treatment in the population by stratifying and matching on the basis the estimated propensity score.

It is found that there is a significant positive effect of being female on job satisfaction which is robust to the different estimation methods applied. Furthermore, this effect does not disappear even when its estimation is based on the most careful matching of male and female employees (radius method of matching for a radius of $\delta = 0.00001$). Based on this statistical evidence, I interpret the results as estimates of the causal effect of gender on job satsfaction.

A question not addressed in this paper and that naturally arises from these results is how to interpret the existence of a positive causal effect of being female on job satisfaction: why should an identical man and woman in identical jobs report different satisfaction scores? In other words, why would the job satisfaction level of an employee be different had he/she the opposite gender? Although the answer to this question is outside the scope of this paper, it is important to comment briefly on the research done on this matter. Clark's (1997) proposed explanation rests on the presence of relative terms in the well-being function, which have been interpreted as an individual's expectations about his/her job. Hence, women will be more satisfied than men with the same objective characteristics and work values if they expect less than men from their job, where women's lower expectations likely result from the poorer position in the labour marked that they have traditionally held.. The basis for this argument is the finding, widespread in psychology, that individuals tend to evaluate experiences relative to some kind of norm or reference level. Both Clark (1997) and Sanz de Galdeano (2001) provide some evidence supporting this interpretation coming from cross-sectional and panel analyses respectively.

DATA APPENDIX

- Female: Respondent is female.
- Age: Age of respondent at date of interview.
- Health dummies (3): Respondents classify their own health, compared to people of their own age. Categories: excellent; good; fair to very poor. Omitted category: fair to very poor.
- Education dummies (3): "High", Degree, teaching qualification or other higher qualification; "Medium", Nursing qualification, A-levels, O-levels or equivalent; "Low", neither of the above. Omitted category: "Low".

There is a small minority of individuals for whom the highest educational qualification obtained reported varies during the observation period, which could be due to marginal variations in educational attainment. However, given the characteristics of the sample, measurement error is a much more likely event. In fact, some of these individuals report a highest educational qualification obtained that decreases over time, which can only be due to measurement error. For this reason, the education dummies are time invariant by construction in the analysis.

- Marital status dummies (6): Married; Cohabiting; Widowed; Divorced; Separated; Never Married. Omitted category: Never Married.
- Occupational category dummies (3): "High occupational category", Professional, managerial and technical occupations; "Medium occupational category", Skilled non manual and skilled manual occupations;"Low occupational category", Partly skilled and unskilled occupations and armed forces. Omitted category: "Low occupational category".
- Log income: Natural log of usual monthly gross pay from respondent's main job.
- Log hours: Natural log of usual weekly hours (including overtime).
- Establishment size dummies (3): Number of workers at establishment is <25 ("Small"); 25-199 ("Medium"); 200+ ("Big"). Omitted category: 200+ ("Big").
- Second job: Respondent has a second paid job.
- Full time job: Respondent is employed full time (works at least 30 hours per week including both normal and overtime hours).

- Temporary worker: Repondent's current job is seasonal, temporary, casual or a job done under contract or for a fixed period of time.
- Union member: Respondent is a member of his/her workplace union.
- Number of own children under the age of 16 in household dummies (5): One kid; Two kids; Three kids; Four or more kids. Omitted category: No kids.
- Carer: Respondent cares for a handicapped, sick or elderly person in household.
- Partner's usual monthly gross pay dummies (5): "the partner does not work or the individual has no partner"; 0-499; 500-999; 1000-1499; $\geqq$1500. Omitted category: "the partner does not work or the individual has no partner".
- Partner's usual weekly hours (including overtime) dummies (5): "the partner does not work or the individual has no partner"; 0-15; 16-29; 30-39; $\geqq$40. Omitted category: "the partner does not work or the individual has no partner".
- Region dummies (18): Standard British regions plus seven metropolitan areas. Omitted category: Rest of South East.
- Industry dummies (10): Agriculture, forestry and fishing; energy and water supplies; extraction of minerals, manufacture of metals, mineral products and chemicals; metal goods, engineering and vehicles industries; other manufacturing industries; construction; distribution, hotels and catering; transport and communication; banking, finance, insurance, business services and leasing; other services. Omitted category: other services.

STATISTICAL APPENDIX

Proof of Lemma 1, Rosenbaum and Rubin (1983):
First,

$$\Pr(T_i = 1|X_i, p(X_i)) = E[T_i|X_i, p(X_i)] = E[T_i|X_i] = p(X_i)$$

because by definition $p(X_i) = E[T_i|X_i]$. Second,

$$\begin{aligned}\Pr(T_i = 1|p(X_i)) &= E[T_i|p(X_i)] \\ &= E[E[T_i|X_i, p(X_i)]|p(X_i)] = E[p(X_i)|p(X_i)] = p(X_i)\end{aligned}$$

Hence $\Pr(T_i = 1|X_i, p(X_i)) = \Pr(T_i = 1|p(X_i))$ and conditionally on $p(X_i)$ the treatment indicator and T_i and the covariates X_i are independent. *QED*.

Proof of Lemma 2, Rosenbaum and Rubin (1983):

First,

$$\begin{aligned}
\Pr(T_i &= 1|Y_i(1), Y_i(0), p(X_i)) = E[T_i|Y_i(1), Y_i(0), p(X_i)] \\
&= E[E[T_i|Y_i(1), Y_i(0), X_i, p(X_i)]|Y_i(1), Y_i(0), p(X_i)] \\
&= E[p(X_i)|Y_i(1), Y_i(0), p(X_i)]
\end{aligned}$$

Second, as shown in the proof for Lemma 1, $\Pr(T_i = 1|p(X_i)) = p(X_i)$. Hence $\Pr(T_i = 1|Y_i(1), Y_i(0), p(X_i)) = \Pr(T_i = 1|p(X_i))$, and conditionally on $p(X_i)$ the treatment indicator T_i and the potential outcomes $Y_i(1)$ and $Y_i(0)$ are independent. QED.

An algorithm for estimating the propensity score, Dehejia and Wahba (1998)

1. Start with a pasrimonious logit/probit function to estimate the score.

2. Sort the data according to the estimated propensity score (ranking from lowest to highest).

3. Stratify all observations such that the estimated propensity scores within a stratum for treated and control units are close (no significant difference):

(a) Start by dividing observations in five blocks of equal score range (0-0.2,..., 0.8-1)

(b) Test whether the differences in means of the propensity scores across treated and controls units within each block are significantly different from zero.

(c) If the answer is yes, increase the number of blocks and test again.

(d) If the answer is no, go to the next step.

4. Test whether Lemma 2 holds in all blocks for all covariates:

(a) For all covariates, test whether the differences in means across treated and control units within each block are significantly different from zero. If the answer is no, stop.

(b) If covariate j is not balanced for some blocks, divide them into finer blocks and re-evaluate.

(c) If covariate j is not balanced for all blocks, modify the logit by adding interaction terms and/or higher order terms of the covariate j and re-evaluate.

The stratification estimator

Let $q = 1, ...Q$ index the strata defined by the propensity score estimation. Then, the estimated average teatment effect is given by:

$$\widehat{ATE}_S = \sum_{q=1}^{Q} \frac{N_q}{N} \left[\frac{\sum_{i \in T_q} Y_i}{N_q^T} - \frac{\sum_{i \in C_q} Y_i}{N_q^C} \right]$$

where N denotes total number of units, N_q represents the number of units in stratum q, T_q and C_q are the sets of treated and control units in stratum q and N_q^T and N_q^C the corresponding number of units. Assume that treated and control units are independent and that the variance of Y_i is the same within each group. Then, the variance of $\widehat{ATE}_S$ can be computed as:

$$Var\left(\widehat{ATE}_S\right) = \sum_{q=1}^{Q} \frac{N_q}{N^2} \left[\frac{N_q}{N_q^T} Var(Y_i|T_i = 1) + \frac{N_q}{N_q^C} Var(Y_i|T_i = 0) \right]$$

The matching estimators

Let m_{ij} be a dichotomous variable that takes value one if and only if unit i is matched to unit j and value zero otherwise. Formally, $m_{ij} = 1$ if and only if:

(a) $T_i \neq T_j$, that is, if the two units do not have the same treatment status.

(b)

(b.1) For the nearest match method:

$$/p(X_j) - p(X_i)/ \leq /p(X_j) - p(X_k)/, \text{ for all } k \text{ such that } T_k \neq T_j$$

(b.2) For the radius method:

$$/p(X_j) - p(X_i)/ \leq \delta$$

where δ represents the tolerance level chosen by the researcher and $p(X_i)$ is the estimated propensity score for unit i.

Then, the non adjusted estimator based on the nearest match method is given by:

$$\widehat{ATE}_{NM} = \frac{1}{N} \left[\sum_{i \in T} \phi_i Y_i - \sum_{i \in C} \phi_i Y_i \right]$$

where T and C are the sets of treated and control units.

The weight associated to each unit, ϕ_i, can be computed as:

$$\phi_i = 1 + \sum_j \frac{m_{ij}}{\sum_i m_{ij}}$$

As for the non adjusted estimator based on the radius method, it is given by:

$$\widehat{ATE}_{RM} = \frac{1}{N_{RM}} \left[\sum_{i \in T} \theta_i Y_i - \sum_{i \in C} \theta_i Y_i \right]$$

where N_{RM} represents the number of units such that $\sum_j m_{ij} \geq 1$.

The weight associated to each unit, θ_i, can be computed as:

$$\theta_i = \phi_i \text{ if } \sum_j m_{ij} \geq 1$$

$$\theta_i = 0 \text{ otherwise}$$

Assume that treated and control units are independent and that the variance of Y_i is the same within each group. Then, the variances of the two estimators are:

$$Var\left(\widehat{ATE}_{NM}\right) = \frac{1}{N^2}\left[\sum_{i \in T}\left[\phi_i^2 \cdot Var(Y_i|T_i = 1)\right] + \sum_{i \in C}\left[\phi_i^2 \cdot Var(Y_i|T_i = 0)\right]\right]$$

and

$$Var\left(\widehat{ATE}_{RM}\right) = \frac{1}{N_{RM}^2}\left[\sum_{i \in T}\left[\theta_i^2 \cdot Var(Y_i|T_i = 1)\right] + \sum_{i \in C}\left[\theta_i^2 \cdot Var(Y_i|T_i = 0)\right]\right]$$

References

Adams, J. S., (1963), "Toward an understanding of inequity", *Journal of Social Psychology*, Vol. 67, 422-436.

Angrist, J. D. and W. Evans, (1998), "Children and their parents' labor supply: evidence from exogenous variation in family size", *American Economic Review* 88, Vol. 3, 450-477.

Akerlof, G. A., A. K. Rose, and J. L. Yellen, (1988), "Job switching and job satisfaction in the US Labor Market", *Brooking papers in Economic Activity*, Vol. 2, 495-582.

Bartel, A. P., (1981), "Race Differences in Job Satisfaction: A Reappraisal", *Journal of Human Resources*, Vol. 16, 295-303.

Bell, D. E., (1982), "Regret in decision making under uncertainty", *Operations Research*, Vol. 30, 961-981.

Bell, D. E., (1985), "Disappointment in decision making under uncertainty", *Operations Research*, Vol. 33, 1-27.

Berger, C., C. Olson, and J. Boudreau, (1983), "Effects of unions on job satisfaction: the role of work-related values and perceived rewards", *Organizational Behaviour and Human Performance*, 32, 289-324.

Bertrand, M. and S. Mullainathan, (2001), “Do People Mean What They Say?”, *MIT Working Paper Series*, No. 01-04.

Blanchflower, D. and A. J. Oswald, (1992), “Entrepreneurship, Happiness and Supernormal Returns: Evidence from Britain and the U.S”, *NBER Technical Working Paper Series*, No. 4228.

Blanchflower, D. and A. J. Oswald, (1997), “The rising well being of the young”, *The labour market consequences of technical and structural change, Discussion paper series*, No. 16.

Blau, F. D. and L. M. Kahn, (1992), “The Gender Earnings Gap: Learning from International Comparisons”, *American Economic Review*, Vol. 82, 533-38.

Borjas, G. J., (1979), “Job Satisfaction, Wages and Unions”, *Journal of Human Resources*, Vol. XIV, No. 1, 21-40.

Clark, A. E., (1996), “Job Satisfaction in Britain”, *British Journal of Industrial Relations,* Vol. 34, 189-218.

Clark, A. E., (1997), “Job Satisfaction and Gender: Why are women so Happy at Work?”, *Labour Economics*, Vol. 4, No. 4, December, 341-372.

Clark, A. E., Y. Georgellis and P. Sanfey, (1998), “Job Satisfaction, Wage Changes and Quits: Evidence from Germany”, *Research in Labor Economics*, Vol. 17, 95-122.

Clark, A. E. and A. J. Oswald, (1994), “Unhappiness and Unemployment”, *Economic Journal*, Vol. 104, 648-59.

Clark, A. E. and A. J. Oswald, (1996), “Satisfaction and Comparison Income”, *Journal of Public Economics*, Vol. 61, No. 3, September, 359-382.

Clark, A. E., A. J. Oswald, and P. B. Warr, (1996), “Is job satisfaction U-shaped in age?”, *Journal of Occupational and Organizational Psychology* 69, 57-81.

Clegg, C. W., (1983), “Psychology of Employee Lateness, Absence and Turnover: A Methodological Critique and an Empirical Study”, *Journal of Applied Psychology*, Vol. 68, 88-101.

Dehejia, R. H. and S. Wahba, (1998), “Propensity Score Matching Methods for Non-experimental Causal Studies”, *NBER Technical Working Paper Series*, No.6829.

Dehejia, R. H. and S. Wahba, (1999), "Causal Effects in Non-experimental Studies: Reevaluating the Evaluation of Training Programmes", *Journal of the American Statistical Association*, Vol. 94, 1053-63.

Freeman, R. B. (1978), "Job satisfaction as an economic variable", *American Economic Review*, May, Vol. 68, 135-141.

Greene, W., 1993. *Econometric Analysis*, 2nd ed., Macmillan, New York.

Gronau, R., (1974), "Wage Comparisons-A Selectivity Bias", *Journal of Political Economy*, Vol. 82, 1119-43.

Hamermesh, D. S., (1977), "Economic aspects of job satisfaction", in Ashenfelter, O. E. and Oates, W. E., editors, *Essays in Labor Market Analysis*, John Wiley, New York, 53-72.

Hamermesh, D. S., (1999), "The Art of Labormetrics", *NBER Technical Working Paper Series*, No. 6927.

Hamermesh D. S., (2001), "The Changing Distribution of Job Satisfaction", *Journal of Human Resources*, Vol. 36, 1-30.

Heckman, J., (1974), "Shadow Prices, Market Wages and Labor Supply", *Econometrica*, Vol. 42, 931-60.

Heckman, J., (1978), "Dummy Endogenous Variables in a Simultaneous Equations System", *Econometrica*, Vol. 46, 931-60.

Heckman, J., (1979), "Sample Selection Bias as a Specification Error", *Econometrica,* Vol. 47, 153-161.

Heckman, J., H. Ichimura and P. Todd, (1997), "Matching as an Econometric Evaluation Estimator: Evidence from an Evaluation of a Job Training Program", *Review of Economic Studies*, Vol. 64, 605-664.

Holland, P. W., (1986), "Statistics and Causal Inference", *Journal of the American Statistical Association*, Vol. 81, 945-970.

Homans, G. C., 1961. *Social behavior: its elementary forms*, Harcourt Brace, New York.

Idson, T. L., (1990), "Establishment Size, Job Satisfaction and Structure of Work", *Applied Economics*, Vol. 11, No. 4, 606-28.

Imbens, G. W., D. B. Rubin and B. Sacerdote, (1999), "Estimating the Effect of Unearned Income on Labor Supply, Earnings, Savings and Consumption: Evidence from a Survey of Lottery Players", *NBER Technical Working Paper Series*, No.7001.

Killingsworth, J. H. and J. Heckman, (1986), "Female Labour Supply: a Survey", in Ashenfelter Layard, (eds.), *Handbook of Labour Economics*, Vol. 1, 103-204.

Lalonde, R., (1986), "Evaluating the Econometric Evaluations of Training Programmes with Experimental Data", *American Economic Review*, Vol. 76, 604-620.

Lazear, E. P. and S. Rosen, (1990), "Male-Female Differentials in Job Ladders", *Journal of Labour Economics*, Vol. 8, S106-S123.

Mangione, T. W. and R. P. Quinn, (1975), "Job satisfaction, Counterproductive Behaviour and Drug use at work", *Journal of Applied Psychology*, Vol. 60, 114-16.

Manski, C. F., 1995. *Identification Problems in the Social Sciences*, Harvard University Press, Cambridge, Massachusetts.

McEvoy, G. M. and W. F. Cascio, (1985), "Strategies for Reducing Employee Turnover; A Meta-Analysis", *Journal of Applied Psychology*, Vol. 70, 342-53.

Meng, R., (1990), "The Relationship between Unions and Job Satisfaction", *Applied Economics*, Vol. 22, 1635-48.

Miller, P. W., (1990), "Trade Unions and Job Satisfaction", *Australian Economic Papers*, Vol. 29, 226-248.

Mincer, J., (1962), "Labor Force Participation of Married Women: A Study of Labor Supply", *International Library of Critical Writings in Economics*, Vol. 45, 211-53. Aldershot, U.K.

Mincer, J., (1963), "Market Prices, Opportunity Costs, and Income Effects", *Collected essays of Jacob Mincer*, Vol. 2. Studies in labor supply. Economists of the Twentieth Century Series. Aldershot, U.K.: Elgar; distributed in the U.S. by Ashgate, Brookfield, 36-52.

Palmore, E., (1969), "Predicting longevity: a follow-up controlling for age", *Journal of Gerontology* 39, 109-116.

Persson, T., G. Tabellini and F. Trebbi, (2001), "Electoral Rules and Corruption", *CEPR Discussion Paper Series*, No. 2741.

Riach, P. A. and J. Rich, (1987), “Testing for Sexual Discrimination in the Labour Market”, *Australian Economic Papers*, Vol. 26, 165-178.

Rosenbaum, P. R. and D. B. Rubin, (1983), ”The Central Role of the Propensity Score in Observational Studies for Causal Effects”, *Biometrika*, Vol. 70, 41-55.

Rubin, D., (1974), “Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies”, *Journal of Educational Psychology*, Vol. 66, 688-701.

Sales, S. M. and J. House, (1971), “Job satisfaction as a possible risk factor in coronary heart disease”, *Journal of Chronic Diseases* 23, 861-873.

Sloane, P. J. and H. Williams, (2000), “Job Satisfaction, Comparison Earnings and Gender”, *Labour*, Vol. 14, No. 3, 473-501.

Sloane, P.J. and M.E. Ward, (2000),“Non Pecuniary Advantages versus Pecuniary Disadvantages: Job Satisfaction among Male and Female Academics in Scottish Universities”, *Scottish Journal of Political Economy,* Vol. 47, No*. 3, 273-303.*

Sloane, P.J. and K. A. Bender, (1998), “Job Satisfaction, Trade Unions, and Exit-Voice Revisited”, *Industrial and Labor Relations Review*, Vol. 51, No. 2, 222-40.

Weiler, W. (1990), “Integrating Rank Differences into a Model of Male-Female Salary Discrimination”, *Quarterly Review of Economics and Business*, Vol. 30, No. 1, 3-15.

Wright, R. E. and P. J. Ermish, (1991), “Gender Discrimination in the British Labour Market: A Re-Assessment”, *Economic Journal*, Vol. 101, No. 406, 508-521.

Zavoina, R. and W. McKelvey, (1975), “A Statistical Model for the Analysis of Ordinal-Level Dependent Variables”, *Journal of Mathematical Sociology*, Summer, 103-120.

Table 1: Time Structure of the Data

N. Obs.	**Year**								
	1991	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**Total**
All	4,421	4,145	3,971	4,053	4,007	4,179	4,293	4,788	33,857
Men	2,247	2,079	1,952	2,020	2,005	2,071	2,172	2,388	16,934
Women	2,174	2,066	2,019	2,033	2,002	2,108	2,121	2,400	16,923

Note: for each year the table reports the number of non-missing observations on all the variables used in the statistical analyses.

Table 2: Sample Means of Job Satisfaction by Gender

	Mean	**S.D.**	**Min.**	**Max.**	**t-statistic**		
All	5.38	1.36	1	7			
Men	5.21	1.39	1	7	-23.88		
Women	5.56	1.31	1	7			
Scale Response	**All**	**%**	**Men**	**%**	**Women**	**%**	**t-statistic**
1	682	2.01	388	2.29	294	1.74	3.62
2	900	2.66	533	3.15	367	2.17	5.60
3	2,225	6.57	1,333	7.87	892	5.27	9.66
4	2,908	8.59	1,813	10.71	1,095	6.47	13.94
5	7,101	20.97	3,891	22.98	3,210	19.97	9.07
6	14,103	41.65	6,709	39.62	7,394	43.69	-7.60
7	5,938	17.54	2,267	13.39	3,671	21.69	-20.21
Total	33,857	100.00	16,934	100.00	16,923	100.00	

Note: statistics based on the 33,857 employee-year observations for which complete information is available on all the variables used in the statistical analyses.

Table 3: Sample Means of Characteristics by Gender

Variable	All		Men		Women		t-statistic
	Mean	S.D.	Mean	S.D.	Mean	S.D.	
Age	36.95	11.62	36.86	11.62	37.03	11.62	-1.32
Excellent health	0.28	0.45	0.31	0.46	0.26	0.43	10.64
Good health	0.50	0.49	0.49	0.50	0.51	0.49	-2.81
High education	0.33	0.47	0.38	0.48	0.29	0.45	17.62
Medium education	0.44	0.49	0.41	0.49	0.48	0.49	-12.76
Married	0.58	0.49	0.59	0.49	0.57	0.49	3.58
Cohabiting	0.12	0.32	0.11	0.32	0.12	0.32	-1.52
Divorced	0.04	0.21	0.03	0.17	0.06	0.24	-14.33
Separated	0.01	0.12	0.01	0.10	0.02	0.14	-7.60
Widowed	0.01	0.11	0.004	0.06	0.02	0.14	-13.63
Log income	6.74	0.80	7.09	0.61	6.39	0.82	87.27
Log hours	3.54	0.51	3.75	0.32	3.34	0.57	82.45
Small	0.33	0.47	0.28	0.45	0.38	0.48	-19.82
Medium	0.37	0.48	0.37	0.48	0.36	0.48	3.08
High occupational category	0.34	0.47	0.38	0.48	0.31	0.46	13.61
Medium occupational category	0.46	0.49	0.44	0.49	0.47	0.49	-6.03
Second job	0.10	0.30	0.09	0.29	0.11	0.31	-6.15
Full time job	0.79	0.40	0.95	0.20	0.63	0.48	79.34
Temporary worker	0.08	0.27	0.06	0.25	0.09	0.29	-8.32
Union member	0.22	0.41	0.23	0.42	0.21	0.41	5.17
One kid	0.15	0.36	0.14	0.35	0.16	0.36	-4.12
Two kids	0.14	0.35	0.14	0.35	0.14	0.34	1.66
Three kids	0.04	0.20	0.04	0.20	0.03	0.19	3.17
Four or more kids	0.008	0.08	0.009	0.09	0.006	0.08	3.37
Carer	0.02	0.16	0.02	0.16	0.02	0.16	-1.00
Partner's income 0-499	0.10	0.31	0.18	0.39	0.02	0.16	49.38
Partner's income 500-999	0.16	0.36	0.19	0.39	0.13	0.34	14.25
Partner's income 1000-1499	0.14	0.34	0.09	0.28	0.19	0.39	-26.71
Partner's income $\geqq$ 1500	0.15	0.36	0.06	0.24	0.25	0.43	-49.47
Partner's hours 0-15	0.04	0.19	0.07	0.26	0.006	0.07	32.21
Partner's hours 16-29	0.07	0.26	0.13	0.34	0.01	0.10	45.31
Partner's hours 30-39	0.15	0.36	0.18	0.38	0.13	0.34	11.62
Partner's hours $\geqq$ 40	0.29	0.45	0.14	0.34	0.45	0.49	-67.28

Note: Statistics based on the 33,857 employee-year observations for which complete information is available on all the variables used in the statistical analyses.

Table 4: Parametric Estimates of the Effect of Gender on Overall Job Satisfaction

Single Equation Models			**Heckman Selection Model**[C]		
(1)	(2)	(3)	(4)	(5)	(6)
OLS[A]	**Ordered Probit**[B]	**N. Obs.**	**Full MLE**	**2-STEP**	**N. Obs.**
0.226	0.179	36,100	0.208	0.195	33,857
(0.016)	(0.013)		(0.017)	(0.017)	

Notes:

Coefficients on the Female dummy variable are reported. Standard errors in parentheses.

(A) OLS regression: Job satisfaction on a constant, female dummy, age, age-squared/1000, education dummies, health dummies, marital status dummies, establishment size dummies, occupational category dummies, log income, log hours, second job, full time job, temporary worker, union member, region dummies and industry dummies.

(B) Ordered probit specification is the same as in note (A).

(C) The control variables included in the structural equation are the same as those entering the previous OLS and ordered probit regressions. The selection equation is estimated using a probit model for being an employee that includes all of the non-job control variables previously used plus a number of other variables that are excluded from the structural equation. They include partner's pay, partner's hours of work, the provision of care for others and the number of own children in the household, all of them interacted with gender.

Table 5: Propensity Score Estimates of the Effect of Gender on Overall Job Satisfaction

	Unadjusted	**Adjusted**[A]		**N. Obs.**
		OLS	**Ordered Probit**	
	(1)	(2)	(3)	(4)
(1) **Stratification Estimate**	0.261	0.176	0.135	33,857
	(0.033)	(0.033)	(0.027)	
Matching Estimates:				
(2) **Nearest Match**	0.289	0.284	0.236	33,857
	(0.010)	(0.014)	(0.011)	
(3) **Radius** $\delta = 0.0001$	0.258	0.249	0.201	21,903
	(0.011)	(0.018)	(0.014)	
(4) **Radius** $\delta = 0.00005$	0.250	0.246	0.202	15,759
	(0.014)	(0.021)	(0.016)	
(5) **Radius** $\delta = 0.00001$	0.243	0.243	0.207	4,924
	(0.027)	(0.038)	(0.030)	

Notes:

Coefficients on the Female dummy variable are reported. Standard errors in parentheses.

The propensity score is estimated using a probit model with the following specification:

$\Pr(\text{Female dummy} = 1) = F($age, age-squared/1000, education dummies, health dummies, marital status dummies, establishment size dummies, occupational category dummies, log income, log hours, second job, full time job, temporary worker, union member, region dummies, industry dummies, number of own children in household dummies, carer, partner's pay dummies and partner's weekly hours of work dummies).

(A) Estimation by regression adjustment controls for all covariates linearly. For matching with replacement, weighted regression is used, where each treated (control) unit's weight depends on the number of times it it is matched to a control (treated) unit. The exact expression for the weights is given in the Statistical Appendix.

Table 6: Propensity Score Estimates of the Effect of Gender on Job Satisfaction with Total Pay

	Unadjusted	Adjusted^{A}		N. Obs.
		OLS	Ordered Probit	
	(1)	(2)	(3)	(4)
(1) **Stratification Estimate**	0.409	0.275	0.192	33,838
	(0.039)	(0.040)	(0.026)	
Matching Estimates:				
(2) **Nearest Match**	0.374	0.312	0.213	33,838
	(0.012)	(0.017)	(0.011)	
(3) **Radius** $\delta = 0.0001$	0.347	0.326	0.213	21,895
	(0.014)	(0.021)	(0.014)	
(4) **Radius** $\delta = 0.00005$	0.355	0.341	0.223	15,818
	(0.017)	(0.025)	(0.016)	
(5) **Radius** $\delta = 0.00001$	0.420	0.398	0.272	4,990
	(0.032)	(0.045)	(0.029)	

Table 7: Propensity Score Estimates of the Effect of Gender on Job Satisfaction with the Actual Work Itself

	Unadjusted	Adjusted^{A}		N. Obs.
		OLS	Ordered Probit	
	(1)	(2)	(3)	(4)
(1) **Stratification Estimate**	0.261	0.176	0.135	33,861
	(0.033)	(0.033)	(0.027)	
Matching Estimates:				
(2) **Nearest Match**	0.211	0.197	0.152	33,861
	(0.010)	(0.015)	(0.011)	
(3) **Radius** $\delta = 0.0001$	0.168	0.161	0.124	22,042
	(0.012)	(0.018)	(0.014)	
(4) **Radius** $\delta = 0.00005$	0.185	0.176	0.138	15,888
	(0.014)	(0.021)	(0.016)	
(5) **Radius** $\delta = 0.00001$	0.227	0.221	0.172	4,997
	(0.027)	(0.039)	(0.030)	

Note: see footnotes on Table 5.

Figure 1: Histogram of the Estimated Propensity Score for BHPS Treated and Control Units