DISCUSSION PAPER SERIES

No. 4223

# THE EFFECT OF DISABILITY ON LABOUR MARKET OUTCOMES IN GERMANY: EVIDENCE FROM MATCHING

Michael Lechner and
Rosalia Vazquez-Alvarez

*LABOUR ECONOMICS*

Centre for Economic Policy Research

www.cepr.org

Available online at: www.cepr.org/pubs/dps/DP4223.asp

ISSN 0265-8003

# THE EFFECT OF DISABILITY ON LABOUR MARKET OUTCOMES IN GERMANY: EVIDENCE FROM MATCHING

**Michael Lechner,** University of St Gallen and CEPR
**Rosalia Vazquez-Alvarez,** University of St Gallen

Discussion Paper No. 4223
February 2004

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 4223

February 2004

# ABSTRACT

## The Effect of Disability on Labour Market Outcomes in Germany: Evidence from Matching

If labour market policies aimed at people with disabilities are effective, we should observe no significant difference in labour market outcomes between disabled and non-disabled individuals. This Paper examines the impact of disability status on labour market outcomes using matching methods associated with treatment effect techniques for programme evaluation. Such techniques are fairly robust with respect to model misspecification and account for the common support problem, thus improving the identification and estimation strategy. Using the German Socio-Economic Panel (1984-2001) we estimate the impact of disability on labour market participation and different income measures. We find that those who are not disabled experience higher employment rates and higher earnings relative to those who have become disabled. This difference is almost always significant for all labour market outcomes considered.

JEL Classification: C13, C14, I12, I18 and J23
Keywords: causality, evaluation of disability policies, health status, labour market outcomes, matching on the propensity score and treatment effect

Michael Lechner
SIAW
University of St Gallen
Dufourstrasse 48
St Gallen 9000
SWITZERLAND
Tel: (41 71) 224 2350
Fax: (41 71) 224 2298
Email: michael.lechner@unisg.ch

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=139325

Rosalia Vazquez-Alvarez
SIAW
University of St Gallen
Dufourstrasse 48
St Gallen 9000
SWITZERLAND
Email: rosalia.vazquez-alvarez@unisg.ch

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=160133

Submitted 05 December 2003

# 1 Introduction

Most industrialized economies recognize the need for effective policies and practices in support of workers whose prospects of either remaining or (re-)integrating in employment are jeopardised by work injury, long term illness and/ or disability. For example, in Germany the Severely Disabled Persons Act of 1974 (*Schewerbehindertengestz Schwbg*) – further amended in 1986 -, or SPDA for short, sets forth the obligation of an statutory quota of a minimum of 6% disable employees on employers with a workforce of 16 employees or more. The same Act obligates the employer to adjust their premises in order to accommodate disable workers, provides legislation which protects the disable against discrimination in recruitment, employment, and unfair dismissal, as well as setting down fines for those who fail to fulfil their quotas, along side a variety of generous subsidies to facilitate employers to adjust to such policies and practices. Likewise, the SDPA provides a wide range of advantages to encourage participation in paid labour market activities of disable individuals who are able to participate, for example, tax benefits, subsidized transport costs, re-training programs and the legal right to longer holidays per year, among others. Countries such as the UK, the USA, and Australia, follow practices similar to those in Germany.[1]

Research focusing on the effect of disability on labour market outcomes is in a sense very similar to empirical studies which focus on labour market outcomes differentials between genders or due to racial differences. Nevertheless, studies of the effect of disability on labour market outcomes is by no means as prolific, specially in Europe. In the United States many studies have focused on the importance of health status (i.e., disability status) on labour supply behaviour, but have centred attention on the population nearing retirement age (for example, see Kreider and Pepper (2002) and references therein, and

[1] For example, in the UK, the 1944 Disable Persons Employment Act (further amended in 1996), imposes an statutory quota of 3% of disable persons in the workforce for employers with 20 or more workers, imposing fines on those whose quotas are not met. The same Act defines the obligations on behalf of employers to adjust their premises in order to accommodate disable individuals, as well as legislation for the protection of disable employees with respect to discrimination in recruitment, employment, or dismissal for reasons which relates to disability.

Williamson and McNamara (2002)). With respect to Europe, examples of studies of the effect of health on labour outcomes are those of Sundberg (1996) – where health is self-reported health –, Walker and Thompson (1996), and Kidd, Sloane and Ferko (1998) – both of which allow define health with respect to disability levels. In the study by Walker and Thompson (1996), the selection process into employment is explicitly accounted for in a model that estimates the effect of various measures of disability on both hourly wages and labour force participation using British longitudinal data on males. They find that a health status that implies disability reduces wages, although it has an even stronger effect on the probability of participating into paid labour market activities. In fact, once they endogeneity of schooling on health is accounted for, the effect of disability status on wages is very small. The Kidd et al. (1998) study also provide an examples on the effect of disability on both wages and participation rates using data on males from the UK 1996 Labour Force Survey. Their study estimates the participation rates of disable on non-disable individuals using independent probit models for each of the two sub-populations. Following Even and McPherson (1991), they decompose the difference between the two estimated participation rates in explained and unexplained components. They find a 50% participation rate differential, and suggest that only half of this estimate can be explained by productivity related characteristics, thus providing evidence on the ineffectiveness of UK labour market policies which aim at integrating the disable into the labour force. Contoyannis (2001) and Aakvik (2003) provide the most recent studies of health status on labour market outcomes in Europe: whereas Contoyannis (2001) still relies on fairly strong parametric assumptions to identify the effects of (self-reported general and psychological) health on hourly wages, Aakvik (2003) offers a more flexible framework to analyse the labour force participation rate on a sample of (objectively diagnosed) disable individuals, conditional on the event that they have received the impact of particular labour market policies. In his study, Aakvik (2003) finds that disable individuals who have received the benefits of further educational programs are at least 5% more likely to find employment relative to those who are also diagnosed as disable but do not receive an increase in their existing level of general educational achievement.

Other than in the Aakvik (2003) study, the key econometric difficulty in the aforementioned literature results from the nonrandom selection of individuals into different status with respect to disability, i.e., workers in sectors with higher occupational hazard, individuals with at taste for sports with high risk, living on a highly urbanized metropolitan area, are all factors which increase the chances of an individual to become disable. For example, in Kidd et al (1998) identification of the effect of disability in the presence of such nonrandom selection comes from conditioning on pre-determined observed characteristics of both participants and non-participants, assuming that such observable characteristics will account for any bias which might result from differentials in the chances of becoming disable.

There are two potential problems with this approach. The first problem is that for any given set of conditioning variables, we might fail to observe persons in each of the two states (disable and non-disable) we seek to compare, known as the failure of the common support condition.

The second problem is that even when the support problem is not an impediment in identifying disability effects, the choice of model (parametric, semiparametric) is often based on strong functional form assumptions, to the extend that model misspecification might also lead to a second source of bias. For example, in the Contoyannis et al. (2001) study, their results are subject to the rather strong assumption that health status is uncorrelated with time variant covariates and/or individual-specific error that might determine the outcome wages, an assumption that the study itself puts in question in the light of the finding in Sundberg (1996).

The two above mention problems are typical when evaluating the effect of a particular treatment (in our case disability) using non-experimental data. However, recently, micro-econometricians have adopted the techniques of epidemiologists based on studying the effect of an intervention (or treatment) to evaluate non-experimental rather than laboratory data.[2] By using nonparametric techniques, such as matching procedures, it is possible to address both the common support issue and problems associated with model specification. As with other commonly used specifications, matching also assumes selection on observable characteristics. The idea is that there exists a set of observed variables such that conditional on these, the impact of the treatment is independent of the outcome that would occur without disability. Such assumption is known as the Conditional Independence Assumption (CIA). Using matching methods requires us to assume that given a set of $X$ variables becoming disable is unrelated to what various definitions of labour market outcomes would be if she had not become disable. Thus, conditional on $X$, we can find a counterfactual outcome to each treated observation and estimate the impact of the treatment.

The empirical results of this paper are based on data from the sample of West German respondents to the German Socio Economic Panel (GSOEP, 1984-2001).[3] This annual survey is very informative with respect to labour market outcomes as well as on social, economic and living conditions in Germany. The panel dates from 1984, and starting from this first wave a section on health issues elicits both disability status as well as degree of disability from each surveyed individuals. We believe that the richness of the data allows us to make the assumption that outcomes (labour market participation, earnings, income) and disability status are independent conditional on observed attributes, thus solving the identification problem inherent in causal analysis.

Besides our contribution to a growing body of applied econometric literature using treatment effect techniques for program evaluation, this paper wants to contribute to the understanding of how disable individuals fair in the labour market. Our empirical results indicate the following conclusions. We have examined the impact of disability labour policies on the labour market participation rate, annual labour earnings and per capita household disposable income. We consider two different sample selection criteria. In the first selection (SSC1), the underlying population are prime age individuals irrespective of their

---

[2] See Angrist (1991), Heckman and Horz (1989), Ichimura and Todd (1997), Lechner (1995, 1996, 1997), Smith and Todd (2000), but to mention a prominent few.

[3] There are various reasons for selecting the West German sample only. One reason is the need for many waves in order to construct the groups of control and treatment (see Section 4), and whereas West Germany has been providing information to the panel since 1984, information from East Germany is only available from 1990. Furthermore, disability policies might have had different effects in these two regions, as it might take various years after unification for East German labour markets to react to such policies in the same way as West German labour markets do.

working status at the beginning of the sequence under observation, while the second selection criteria (SSC2) allows for an underlying population of individuals who, been at prime age with respect to labour market participation, declare to be employed on a full time basis before they might become disable. In both cases our results suggest that those who receive the impact of disability are less likely to be working, will see their labour income reduced and will have lower per capita disposable income, relative to those individuals who, other than not having received the impact of becoming disable at the beginning of the period, are nevertheless identical to the disable population. For example, allowing for SSC1, estimates of the mean impact of disability on the disable show that non-disable are between 3% and 10% more likely to be in employment relative to prime age individuals who declare a disability status. This impact is almost always statistically significant. Analogous estimates for SSC2 show that such difference is between 8% and 12%, and is also statistically significant. With respect to labour income, our estimates for SSC1 show that mean annual labour earnings for the non-disable can be as much as DM 6,200 (DM 10,700 for SSC2) higher than for those who, other than been disable, are identical to their non-disable counterparts. Not only is this difference statistically significant, but it also represents an earning gap of approximately 16% (20%). The difference with respect to per capita disposable household income – thus allowing for a measure of purchasing power inequality – shows lower differentials, with at most a gap of DM 2,500. However, as with all other outcomes considered, the difference is significant, with implications for social policies for the disable, since per capita household income is a measure that takes into account both household composition as well as overall government policy because it directly reflects government intervention in the form of benefits and transfers between groups.

In analysing our results, we make the very plausible assumption that, for each of the sequences of three years, the policies have had time to work on the outcomes of those whose status is disability (see Aakvik (2003) for a comparative time span). Thus, overall, our results suggest that in Germany, the impact of disability policies on the disable are not effective at reducing their participation cost into competitive labour market activities. We analyse 18 years and our results appear to be fairly similar over time.

The remainder of this paper is organized as follows. Section 2 explains the labour market policies and legislation in Germany with respect to disable persons. Section 3 describes the GSOEP data used in the empirical section. Section 4 defines the econometric methodology, identifying conditions and matching methods used. Section 5 presents the estimated impact of disability on different labour market outcomes. Section 6 concludes.

# 2 Policies and practices in Germany for labor market participants with disabilities

The main legislation concerning disable persons in Germany is the "Severely Disabled Persons Act (1974)" – *Schwerbehindertengesetz*– which was further amended in 1986 [4] and issued by the Federal Ministry for Labour and Social Affairs. In short, we refer to this Act as the SDPA. Although the SDPA does not adhere to one exact definition of disability, in its broader terms it takes up the three tiered definition proposed by the World Health Organization (WHO), where disable persons are defined as those who suffer from the consequences of the effects of a physical, mental or psychological condition which is not typical for the respective age, and where the consequences are not merely of a temporary nature. The definition covers the terms handicap, disability and impairment.[5] With such definition as a benchmark, each individual who wishes – voluntarily – to be assessed in terms of disability has to go through a formal medical procedure conducted by a special independent institution (*Versorgungsamt*), where he or she is identified with a particular degree of disability. The degree of disability is express in percentage increments from 0 to 100% (total disability). The degree of disability is given to each person independently from his or her fitness to work in his or her present occupation or in future view of desired occupation. Once an individual is assigned a particular degree of disability, the public welfare authorities (*Hauptfuersorgestellen*) decides if the legislation as set in the SDPA is applicable to that person. Two

[4] See Footnote 1.

possibilities exist. First, legislation as set in the Act covers all individuals with a degree of disability greater or equal to 50%. Second, individuals with a degree between 30 and 50% are also covered if the *Hauptfuersorgestellen* considers that the disability is the reason why the individual cannot find or hold an existing job. The SPDA prescribes and legislates for both sides of the labour market, namely the employer and the employee. Whereas the SDPA provides legislation, prescriptions, penalties and benefits for the employer, legislation with respect to employees are penalty free and only with the voluntary consent of the disable person.

The SPDA legislates that employers with a workforce greater or equal to 16 are legally obliged to employ a minimum of 6% disable workers. Furthermore, employers subject to the legislation have to provide adequate workspace for disable employees, according to their skills and capabilities, as well as appointing a representative inside the workplace who will look after the disable person's interest. Employers who do not fulfil the quota have to pay a levy of DM 200 (equivalent to approximately €105) per month for unfulfilled compulsory placements.[6] This revenue is used fully to finance national measures for the integration of severely disable persons. Since the quotas system was introduced in 1974, the fulfilment of the quota has steadily declined over the years; while the 6% target has never been achieved, the highest percentage was in 1982 with 5.9% with the latest figures showing an average for West Germany of 4.2% over the period since the Act was passed (Zentras, 1997).[7] One could think of such figures as a measure that the policies are not working, and consequently disable are less likely to be employed than non-disable. However, other evidence suggest that what such figures show is a badly

[5] The definition varies according to additional requirements for the application to specific situations, and with regards to the assistance required by different circumstances and institutions (*Bunderministerium fuer Arbeit und Socialordnung (BMA, 1996 Publication, p.11)*)

[6] An alternative to paying the full levy, enterprises can see their levy reduced if they award contracts to sheltered workshops. This workshops are places where severely disable individuals participate on paid labour market activities while sheltered from the competitiveness of the labour market. It is often the case that mentally handicap individuals, e.g., Down Syndrome persons, will work in such shelters.

[7] In general the public sector is better at fulfilling its quotas – e.g., the federal government has to report to parliament every year on such quotas, so it makes an effort to employ at least up to the minimum of 6%.

designed quotas system. In 1995, and according to the quota requirement, there should have been 397,700 vacancies allocated to the disable in West Germany, but during that year only 155,500 severely disable persons (i.e., with at least 30% degree of disability) were registered as unemployed.[8] In 1999 the figures for West Germany were 513,187 required vacancies versus 181,200 registered unemployed with disabilities, while there was an almost balance with respect to the number of disable and vacancies offered to them with a ration of 112:110 (although this does not indicate the ratio of match vacancies). Furthermore, the quota system does not take into account the number of disable employees who are employed beyond the required quota and companies who, without an obligation, still employ disable individuals (Albrecht and Braun, 1998).

The quota legislation comes along with other financial mechanisms that benefit the employer, with an aim to retain and or (re-)employ disable people. Examples of these are subsidizing the creation of new vacancies that would otherwise not be created, wage subsidies with reference to existing vacancies (this can be up to 80% of gross wages for the first year, 70% for the second year and 60% for the third year), financial support for the adaptation of the workplace (with loans and subsidies of up to 100%) and financial support for special employee training and vocational rehabilitation which can also cover up to 100% of the cost. However, according to Thornton and Lunt (1997) the reason why this financial mechanism are rarely taken up by employers is because of lack of information, specially for small enterprises, as well as too much bureaucratised procedures which discourages many small and medium size employers. Furthermore, in the case where benefits might only extend to workspace adaptation, perhaps this is not needed, at least not for existing employees. One further possibility for the failure of enterprise to take advantage of such benefits is because of the double role of the Hauptfuersorgestellen; while employers might take up some of the benefits, they also become fully subject to the sovereignty of the authorities, and this might make employers to be reserved (Albrecht and Braun, 1998). On the other hand,

[8] Although East Germany is not the target of our study, is worth noting that the same quota problem applies to this region of Unified Germany, since figures for East Germany show 20,000 disable persons registered as unemployed, versus 107,000 quota required vacancies.

it is often the case that such subsidies might end up having a dead weight effect with respect to promoting additional disability employment, since employers who receive the subsidy might have employed (or continued to employ) the disable individual anyway. Already in the late 1980's (Oyen, 1989), in Germany, it was noticed that financial incentives do not promote and/ or maintain employment of disable people, but rather they reinforce a willingness to do so for the already existing disable workforce.

The SDPA also sets legislation for the protection of disable employees making dismissal of such workers a very difficult task. if an employer decides to dismiss a disable individual, the representative of the disable in the workplace has to be informed, and such dismissal has to be approved by the welfare authorities (*Hauptfuersorgestellen*). Such protective measures apply also to individuals whose disability degree is been ascertained (e.g., those who become disable with respect to or outside work, are given protection as if severely disable, at least until their disability degree is been assessed). The decision of the *Hauptfuersorgestellen* is mandatory, unless there is some outside agreement on behalf of the employer and employee which satisfies both parties. The basic guidance is that the dismissal will be approved if the employer can proof that the employee stands against the interest of the enterprise. If the dismissal is not approved the employer can appeal to a labour court. In 1995, 35% of such dismissals resulted in job retention (with 15.5% been in disapproval with the employer) while 46.9% resulted in job loss without the consent of the disable employee. The reminder (18.1%) also resulted in job loss but with consent of the employee (e.g., early retirement).

Besides legal protection, disable are also offered financial incentives to encourage them into paid working activities. These include financial support of vocational rehabilitation measures, reimbursement of the cost resulting from job search activities (e.g., application forms, travelling expenses), financial assistance to set up self-employment, purchase of working aids, subsidizing public and private transport, and subsidizing expenses associated with promoting mobility (e.g., subsidize adaptation of a new house if reallocating for work reasons).

All the above legislations and prescriptions should motivate profit maximising employers to employ a percentage of disable at least up to the minimum quota. Likewise, such policies should increase the motivation of disable persons who are capable to enter a competitive labour market, since the aim of such policies is to lower the entry cost of participation.

Overall, if such policies work, we should observed no differential between disable and non-disable participation rate. Wages subsidizing and tax incentives should also account for disability related productivity differentials. Social scientist suggest otherwise, and focus attention on macroeconomic figures as a way to back up the argument that persons with disability fair worst in the labour market than non-disable. A set of figures often mentioned is the overall unemployment rate. For example, Albrecht and Braun (1998) compare the 1996 unemployment rate of officially unemployed disable persons in West Germany (15.9%) to that of the non-disable population (9.1%), and suggest this figures as evidence that the policies do not work. However, this figures compares groups without telling us about the causal relation between disability status and employment status. It might be that the disable who are registered unemployed are associated with occupational sectors that suffers from higher unemployment rates than the non-disable in the population, thus the above snap-shot provides a distorted comparison between two sub-populations. Table 1 shows the distribution of disable employees among economic sectors using data for two selected years (1995 and 2000).[9] For any of the two years, disable individuals are more likely to be associated with blue collar occupations (manufacturing, transportation, production and related) than any of the other economic sectors. Furthermore, in 1995 the difference between disable and non-disable blue collar workers in terms of percentage is positive (disable are 7.2% more likely to belong to this sector) and significant (the estimated t-value of the difference is 2.1). Comparing estimates between 1995 and 2000

---

[9] The study by Albrecht and Braun (1998) provides a brake-down of disable employees between occupational sectors using the national estimates of 1995 based on Zentras (1997). The problem is that neither Albrecht and Braun (1998) or Zentras (1997) provide comparative figures for non-disable employees. The GSOEP data set used in Table 1 provides a representative sample of the West German population. Using the 1-digit ISCO classification, allows for a brake down occupational sector similar to that in the study of Albrecht and Braun (1998), with the added advantage that we can compare disable to non-disable employees. Our estimates compare very well to those in Albrecht and Braun (1998). For example, in their study the share of disable employees in manufacturing, transportation, building and construction is equal to 45.9%, while the share of such employees in sales and services (trade, banking and insurance) is 21.9%. Our estimates (2000) are 42.3% and 22.6, respectively.

suggests that the share of disable and non-disable between occupational sectors have not changed over time. The other two sectors that employ a relatively large percentage of disable persons are the service sector and that related to office-work (e.g., clerk, bookkeeper, etc), however the difference between disable and non-disable for these two sectors is almost negligible. Nevertheless, occupation by sector differs between disable and non-disable, therefore it is not sufficient to simply compare the overall unemployment rate of West Germany between different disability status, but instead, a better practice is to make inference within cells defined by personal attributes, and not only with respect to occupational sector, but also with respect to other characteristics that might affect the employment probability of both disable and non-disable (e.g., ability, motivation, vacancy matching, etc). Examination of micro-economic survey data over time might provide a more robust set of conclusions.

***Table 1: Distribution of disabled among economic sectors, 1995 and 2000 (population estimates for West Germany)***

| | 1995 | | | 2000 | |
|---|---|---|---|---|---|
| **Occupational Category** | **Non-Disable** | **Disable** | | **Non-Disable** | **Disable** |
| **Sample size** | 3,578 | 202 | | 3,220 | 194 |
| **Professional, technical and related** | 16.2 | 11.4 | | 19.0 | 11.9 |
| **Administration and managerial work** | 5.2 | 3.5 | | 4.2 | 2.6 |
| **Clerical and related office work** | 19.3 | 22.3 | | 19.9 | 20.6 |
| **Sales worker** | 7.7 | 5.9 | | 8.3 | 7.7 |
| **Service Worker** | 17.4 | 16.3 | | 16.8 | 14.9 |
| **Agricultural, animals, forestry, fishery** | 2.0 | 0.5 | | 1.8 | - |
| **Production, manufacturing, transport and related** | 25.0 | 32.2 | | 22.6 | 35.1 |
| **Others** | 7.2 | 7.4 | | 7.4 | 7.2 |

Source: West German Sample as defined in the text; 100% GSOEP 1995, 2000.

Note 1: Table 1 is based on the representative sample from West Germany of disable and non-disable who declare to be active labour market participants at the time of the survey. The percentage of disables in the sample is in line with national estimates; according to the *Mikrozensus* statistical survey the average percentage of disable of the workforce (1998 – 2000) in West Germany was between 4.1 and 4.2 % (Statistisches Bundesamt Deutschland, see www.destatis.de). Our sample of disable in employment in the 1995 GSOEP account for 5.9 % of the working population.

## 3 The GSOEP data

The data used in this study is based on 18 waves of the German Socio Economic Panel (GSOEP, 1984-2001). The GSOEP is an annual microeconomic panel with the first wave starting in 1984. In 1990 the panel was extended to cover the new adhered East German states. The aim of the panel is to provide data for the analysis of social, economic and living conditions in Germany, with data representative of the German population at individual, household and family level. The core questions cover demographics, education, labour market status and labour market history, earnings, housing information, health outcomes, household production and a section on subjective valuations (e.g., satisfaction with work, life, etc.). Apart from the core sample representing the full German population, the panel also contains specific sub-samples representative of minority groups, for example, migration workers (those who are German resident but of Spanish, Turkish, Italian or Yugoslav origin), and immigrants (of any origin) who have settled in Germany since 1984.[10]

Interview are carried out face to face, with each household member age 16 or over counting as an individual observation. Questions referring to household issues are answered by an appointed household representative. In 1994 the survey format changed so that for the first time since unification East and West German households received identical questions harmonized into one single questionnaire. This implies that only from 1994 onwards the questions which objectively identifies if a person is legally classified as disable was identical for both East and West Germany. Clearly, the effect of disability policies will differ between East and West Germany for reasons that might not be easily controlled given the available data (e.g., different cultural response behaviour to the status of disability between East and West Germany, with East German employers taking much longer to adapt to disability laws), while the use of weights in order to join the East and West German samples into one single set of data can be problematic since it might not be the case that weights are built accounting for disability status. One other reason for not mixing the samples is the loss of data from 1984 (starting of the panel) to 1994, since it is only in 1994 when the data set provides full information on disability for households surveyed in East Germany. In order to avoid all these problems our analysis centres on the representative sample from West Germany using data from 1984 to 2001.[11]

[10] For a more detail account of the structure and contents of the panel visit www.diw.de

[11] At the time of this version, this is the latest wave to be released in the GSOEP by the DIW.

The health section of the questionnaire identifies both if a person has been assessed for disability and the degree of disability assigned to the person, if any. Appendix A shows the exact wording for all questions in the survey related to disability and disability status. According to disability laws in Germany (see Section 2), an individual can benefit from policies on disability with respect to labour market outcomes if they are assigned a degree of disability of 50% or greater. However, those with a degree between 30 and 50% also fall within the benefits of the policies, and therefore, in our empirical section we identify individuals as disable if they declare a degree of disability equal or greater than 30%.[12]

Our target population becomes the permanent inhabitants of West Germany, before and after unification. This is a population that has received the effect of the SDPA – thus the effect of disability policies with respect labour market outcomes – consistently since its enactment in 1974. Potentially there are 16,455 (unique) adult respondents in the West German sample since 1984 to 2001. For each wave since 1984, we select these individuals and independently for each year we apply the same selection criteria consistently over the 18 years under study. Our first selection criteria (SSC1) selects individual respondents between 17 and 60 years of age, excluding those in full time education and individuals performing military service. We also exclude individuals who declare a degree of disability greater or equal to 90% since these are individuals often employed in sheltered workshops where competitive labour market forces are absent. The selection criteria leads to an unbalanced panel with 10,995 different individuals over the 18 year period, which includes employees, self-employed, registered unemployed as well as those who been at prime labour market age declare a non-employment status, i.e., either early retired or individuals

[12] The fact that there is a clear cut distinction between those in the 30%-49% group and those in the 50%-100% does not imply that the second group are the only ones to benefit from the policies prescribed by the SDPA. In fact, application of the Act is discretionary and becomes very much dependent on the judgement of the labour officer in the implementing institution. Semlinger (1995) shows that it is sufficient to show *some* kind of disability as permanently reducing the chances of integration into working life to benefit fully from legislation in the SDPA. Intuitively, if an individual *voluntarily* submits to an assessment on the degree of disability, we would expect that he or she is already aware of the benefits, if only because of tax incentives and disability allowances. It is therefore very plausible to assume that anyone who has been diagnosed with a degree of disability of 30% or above is treated equally as anyone with a degree of 50% or above.

who declare house work as main activity. Notice that factors such as inadequate information channels, motivation (e.g., inadequate policies do not provide enough motivation for the disable to participate), etc., might result on non-participant disable persons opting not to register as unemployed. Thus, to avoid selecting on characteristics correlated to the efficacy of the policies on disability/labour, and to avoid conditioning on motivation, we focus on working versus non working status as one possible labour market outcome, rather than employed versus registered unemployed. Two further measures of labour market outcomes considered in the empirical section are individual annual labour earnings and per capita household disposable income. Furthermore, we consider a second sample selection criteria (SSC2) that is a sub-sample of SSC1. As suggested in Section 1, we define our overall sample as the sum of sub-samples of individuals who both comply to the sample selection criteria for each of the years under study while they are observed consecutively over sequences of three years each. Whereas in SSC1 individuals are allowed to be of any working status (full time employed, part time employed, unemployed, etc.), in SSC2 this changes so that in the first year of each of the three-year sequences we select those who claim full time employment in the first year of the sequences of three years. By comparing estimates of SSC1 and SSC2 we are performing a sensitivity analysis of the impact of disability since we compare the impact between the overall prime age population and a section of this population that might be more incline to be aware of the policies that aim at reducing entry cost for disable into paid labour market activities.

### *3.1 Constructing the comparison and treatment groups*

One way to observe the effect of becoming disable on labour market outcomes is to examine the labour market outcomes at time $t$ for individuals who been classified as non-disable at some point in the past, became disable at time $t-s$, where $s$ is sufficiently large elapse of time to justify the adaptation of such individuals to the new health status, the workings of disability policies that help disable back into paid labour market activities, and/or a combination of the two. It is to this aim that we use several waves of the panel. Allowing for one time period $(t)$ to represent one year, we define an individual $i \in n$ (where $n$

is the sample size of both treated and non-treated) as a treatment unit, if such person is non-disable at $t_1$, becomes disable in period $t_2$ and remains classified as disable in period $t_3$. We identify such unit with the mnemonic $ADD_i$ – where $A$ defines non-disability status while $D$ defines disability. Individuals in the $ADD$ group can receive the treatment of the policies (at the latest) in the start of the second time period $t_2$ but potentially also as soon as immediately after they have been surveyed in period $t_1$. It is plausible to assume (with a high degree of certainty) that using data from a yearly survey, if a person has been classified as legally disable at $t_2$, the effect of disability policies on their labour market outcomes can only be evident at period $t_3$ and beyond, since it is then when both the policies and the individual's adaptation to the new status will have had an impact on such outcomes. With annual data we require at least three waves to construct the treatment group. The use of 18 waves from the GSOEP (1984-2001) allows the formation of 16 sequences ($S_1$ to $S_{16}$) of three years each ($t_1, t_2, t_3$).[13] Having more than one sequence increases the number of observed treatment units, thus increasing the precision with which we estimate the impact of disability on labour market outcomes. The control (or untreated) group is defined by individuals who declare themselves as non-disable at $t_1$, $t_2$ and $t_3$ at any given sequence, and therefore do not receive the impact of the policies. We define these control individuals with the mnemonic $AAA$.[14]

[13] Clearly, with four or more waves, it is possible to define treatment groups over more than three time periods, for example, define a treatment group over four years, such that we can analyse the effect of disability on labour market outcomes after three years in disability status. One problem immediately obvious is that an increase in the time periods reduces the number of observations in the treatment groups for each of the sequences, thus increasing the uncertainty in the estimation of the effect of disability. For example, allowing for SSC1 and three time periods, the first sequence of three years $S_1$ (i.e., 1984-1986), allows for 46 observations in the treated group, to be compared to 5056 control units. If instead the treatment is observed over four years (i.e., 1984-1987) attrition implies that 12 of the original 46 treated units are no longer observed, thus reducing the treated sample by almost 30%. Ignoring this drop in observations would imply assuming that attrition and disability are independent events, which is a very strong assumption. Similar decreases are observed for all sequences of time periods considered.

[14] When constructing the control group, we clean each sequence (independently) eliminating individuals who classify as AAA in a sequences but, at the same time, declare a degree of disability between 1 and 29% at any of

Having constructed these two groups ( $AAA$ and $ADD$ ) we perform an appropriate comparison of the labour market outcomes of individuals in $AAA$ versus those in $ADD$ at $t_3$.

Table 2 shows the dynamics of the data, the formation of the sequences and the possible combinations between the treated and untreated samples over time. This table shows that an individual who is a control in $S_j$ for $j = 1,...,16$, can be a control unit as many times as 16, that is, for any given sequence. For example, an individual who is a control in $S_1$ and is further observed as non-disable in 1987, is also counted as a control in $S_2$. However, individuals who are observed as controls in various sequences, count as independent observations for each of the different sequences. That is, for $S_{k,k\in J,J\in[1,16]}$, we are interested at the labour market outcomes at $t_3 \mid S_k$ whereas at $S_{l\neq k,l\in J,J\in[1,16]}$ the labour outcomes that need to be compared are at $t_3 \mid S_l$. For example, at $S_1$ we are interested in comparing labour outcomes of disable and non-disable in 1986, whereas at $S_2$ the outcomes of interest are those observed in 1987. The reason for treating controls in different sequences as mutually independent (of course when computing standard errors their correlation is accounted for) is that macro-economic conditions can change over time, as well as the design of the policies affecting disable (and also non-disable) individuals. To control for any bias resulting of these macroeconomic changes over time, we compare labour market outcomes between groups of control and treatments independently within each sequence. The final estimate of the effect of disability on labour market outcomes is based on the average over all the (sequence based) independent estimates.[15]

Although it is possible for an individual to contribute as control unit at each sequence considered, this is not the case for the units in the sample classified as treated. For example, by construction, a treatment unit in the sequence $S_1$, cannot be a treatment unit in $S_2$. In theory it is possible to observe treatment units $S_j$ further participating as treatment units at $S_{j+3}$ and beyond; for example, an individual

---

the three years that makes the sequence in question. Such individuals could in practice – although very unlikely – be benefiting from some sort of disability policies.

15 See Sections 4 and 5 for further details on the estimation techniques, and Appendix B for the algorithm followed in the estimation process.

who is classified as treatment unit in $S_1$ implies that we observe the sequence ADD over the years 1984, 1985 and 1986. If so, we cannot observe this individual as been non-disable in 1985, so that automatically a treatment unit in $S_1$ cannot be a treatment unit in $S_2$. However, the same individual can be observed as reporting ADDADD between 1984 and 1989, such that he or she might count as a treatment unit in both $S_1$ and $S_4$. Allowing for this possibility would imply that disability diagnosis is reversible, which is not the case according to the definition of the SDPA. What is more likely in this rare cases is that the records for the person are subject to data collection error, with the coding of non-disability in 1987 as a coding error. Our data construction strategy consists on eliminating any 'double-counts' of treatment such that if an individual is defined as a treated unit in $S_j$ she or he will not appear again at any $j+1$ follow up sequence, either as control or treatment unit. It is clear from Table 2 that individuals who do not show a (non-)disability pattern which does not allow for either sequence $ADD$ or $AAA$ – or a proper combination of the two – at least once over the 18 years, will not be used in the final estimate of the effect of disability on the labour market outcomes for disable, even if all those who are included in each of the samples – each defined over a three year period – are used in the estimation process in order to correctly account for uncertainty due to sampling error.[16]

[16] In fact, each sequence of three years defines a 'sample set' of observations $n_1,\ldots,n_{16}$. Any one in this sample set is either a control unit, a treatment unit, or neither. The measure of interest (the average treatment effect on the treated) requires the use of treated and control units. However, anticipating Section 4.1 is worth noting at this point that for our purpose each sequence's population is defined as all individuals who comply with the sample selection criteria over the three years, even if their (non-)disability pattern is neither that of a control or a treatment. When estimating the distribution of the measure of interest (either quantiles or standard error), our naïve bootstrap procedure, which is based on drawing with replacement from the original population, takes into account 'anyone who belongs to the population of individuals over the three years', as opposed to simply re-sampling from the group made of controls and treatments only. Although for a small number of draws this would imply lowering the probability of selecting the correct match from the population, a sufficiently large number of draws leads to an empirical bootstrap distribution of the estimated measure that should be identical to the true distribution of the average treatment effect on the treated. See Section 4.1 for more details.

*Table 2: Definition of the comparison and treatment groups*

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | **TREATMENT SAMPLE: ADD** | | | | | |
| | **1984** | **1985** | **1986** | **1987** | **…** | **1998** | **1999** | **2000** | **2001** |
| **$S_1$ [1984-1986]** | $A(t_1)$ | $D(t_2)$ | $D(t_3)$ | | … | | | | |
| **$S_2$ [1985-1987]** | | $A(t_1)$ | $D(t_2)$ | $D(t_3)$ | … | | | | |
| **…** | | | | | … | | | | |
| **$S_{15}$ [1998-2000]** | | | | | … | $A(t_1)$ | $D(t_2)$ | $D(t_3)$ | |
| **$S_{16}$ [1999-2001]** | | | | | … | | $A(t_1)$ | $D(t_2)$ | $D(t_3)$ |
| | | | | **COMPARISON (untreated) SAMPLE: AAA** | | | | | |
| | **1984** | **1985** | **1986** | **1987** | **…** | **1998** | **1999** | **2000** | **2001** |
| **$S_1$ [1984-1986]** | $A(t_1)$ | $A(t_2)$ | $A(t_3)$ | | … | | | | |
| **$S_2$ [1985-1987]** | | $A(t_1)$ | $A(t_2)$ | $A(t_3)$ | … | | | | |
| **…** | | | | | … | | | | |
| **$S_{15}$ [1998-2000]** | | | | | … | $A(t_1)$ | $A(t_2)$ | $A(t_3)$ | |
| **$S_{16}$ [1999-2001]** | | | | | … | | $A(t_1)$ | $A(t_2)$ | $A(t_3)$ |

Out of the original 10,995 individuals who entered the sample at some point between 1984 and 2001, only 8,358 observations contribute to the formation of either control or treatment units.[17]

[17] As previously suggested, we have defined two sample selection criteria, SSC1 and SSC2, each of equal importance to our analysis. However, in order to facilitate the reading and for presentation purpose, the reminder of this section provides estimates, tables and summary statistics with respect to SSC1 only. All analogous estimates, tables and summary statistics with respect to SSC2 – except for final estimate results in Section 5 – are placed in Appendix C. Both samples are of equal importance to our final conclusions and presentation of SSC1 in the main text is merely the consequence that SSC2 is a sub-sample of the latter. We suggest the reader to read Appendix C alongside the main text.

***Table 3: Distribution, for each wave, between Non-disable and Disable, and within group distribution according to degree of disability (annual sample according to sample selection criteria)***

| | | | | NON-DISABLE (degree of disability 0 - 29%) | | | DISABLE (degree of disability 30% - 89%) | | |
|---|---|---|---|---|---|---|---|---|---|
| **Year** | **New Entries** | **Attrition units** | **Net sample size** | **As % of net sample size** | **With degree of disability = 0** | **Disability 1% - 29%** | **As % of net sample size** | **Disability 30% - 49%** | **Disability 50% - 89%.** |
| **1984** | 7074 | - | **7074** | **93.9 (0.3)** | 99.4 (0.1) | 0.6 (0.1) | **6.1 (0.3)** | 26.9 (2.1) | 73.1 (2.1) |
| **1985** | 432 | 900 | **6606** | **94.3 (0.3)** | 99.5 (0.1) | 0.5 (0.1) | **5.7 (0.3)** | 26.9 (2.3) | 73.1 (2.3) |
| **1986** | 399 | 601 | **6404** | **94.0 (0.3)** | 97.9 (0.2) | 2.1 (0.1) | **6.0 (0.3)** | 23.1 (2.1) | 76.9 (2.1) |
| **1987** | 296 | 505 | **6195** | **94.4 (0.3)** | 99.6 (0.1) | 0.4 (0.1) | **5.6 (0.3)** | 24.9 (2.3) | 75.1 (2.3) |
| **1988** | 243 | 595 | **5843** | **94.6 (0.3)** | 99.5 (0.1) | 0.5 (0.1) | **5.4 (0.3)** | 23.2 (2.4) | 76.8 (2.4) |
| **1989** | 228 | 523 | **5548** | **94.5 (0.3)** | 99.4 (0.1) | 0.6 (0.1) | **5.5 (0.3)** | 27.1 (2.5) | 72.9 (2.5) |
| **1990** | 227 | 315 | **5460** | **93.4 (0.3)** | 98.7 (0.1) | 1.3 (0.1) | **6.6 (0.3)** | 50.8 (2.6) | 49.2 (2.6) |
| **1991** | 204 | 312 | **5352** | **94.5 (0.3)** | 99.4 (0.1) | 0.6 (0.1) | **5.5 (0.3)** | 31.8 (2.7) | 68.2 (2.7) |
| **1992** | 198 | 331 | **5219** | **94.4 (0.3)** | 99.3 (0.1) | 0.7 (0.1) | **5.6 (0.3)** | 33.2 (2.8) | 66.8 (2.8) |
| **1993** | 196 | 198 | **5217** | **94.2 (0.3)** | 99.2 (0.1) | 0.8 (0.1) | **5.8 (0.3)** | 42.5 (2.8) | 57.5 (2.8) |
| **1994** | 210 | 344 | **5083** | **94.1 (0.3)** | 99.2 (0.1) | 0.8 (0.1) | **5.9 (0.3)** | 36.7 (2.8) | 63.3 (2.8) |
| **1995** | 207 | 222 | **5068** | **93.7 (0.3)** | 100 | 0 | **6.3 (0.3)** | 40.2 (2.7) | 59.8 (2.7) |
| **1996** | 203 | 300 | **4971** | **93.6 (0.3)** | 99.0 (0.1) | 1.0 (0.1) | **6.4 (0.3)** | 38.1 (2.7) | 61.9 (2.7) |
| **1997** | 186 | 280 | **4877** | **93.8 (0.3)** | 99.2 (0.1) | 0.8 (0.1) | **6.2 (0.3)** | 41.4 (2.8) | 58.6 (2.8) |
| **1998** | 187 | 380 | **4684** | **93.9 (0.3)** | 99.1 (0.1) | 0.9 (0.1) | **6.1 (0.3)** | 41.6 (2.8) | 58.4 (2.8) |
| **1999** | 192 | 325 | **4551** | **94.1 (0.3)** | 99.1 (0.1) | 0.9 (0.1) | **5.9 (0.3)** | 43.0 (3.0) | 57.0 (0.3) |
| **2000** | 164 | 394 | **4326** | **94.0 (0.4)** | 99.1 (0.1) | 0.9 (0.1) | **6.0 (0.4)** | 41.7 (3.1) | 58.3 (0.3) |
| **2001** | 149 | 314 | **4161** | **94.1 (0.4)** | 99.0 (0.2) | 1.0 (0.2) | **5.9 (0.4)** | 39.8 (3.1) | 60.2 (3.1) |

Table 3 shows the distribution of the 10,995; Column 2 shows the number of new entries per year, Column 3 shows number of attrition units per year and Column 4 shows the net number of individuals who comply with initial sample selection for SSC1 (with respect to age, employment status and degree of disability).[18] Column 4 shows that the sample size decreases over time to almost two thirds in 2001 relative to the total observations in 1984: this is a characteristic of the GSOEP data set, rather than a result of our sample selection criteria. The distribution between disable and non-disable is consistent over time, with each of

[18] Notice that sample statistics shown in Table 3 do not as yet show the samples of control and treated group. For example, Table 3 Columns 5 and 8 show that the distribution of the 1984 sample between non-disable and disable is 93.9 and 6.1 percent, respectively. On the other hand, the number of controls and treatments within a sequence is defined over three years. For example, $S_1$ is defined over the years 1984 to 1986, and will combine the patters of (non-)disability of the 7,074 observations in 1984 with that of the 6,606 in 1985 and the 6,404 in 1986 (many of which might, of course, be the same overlapping individual over time). Whereas this latter can be thought as the dynamic process of disability over time – see Table 6 for this purpose – Table 3 shows the distribution of non-disables versus disable individuals as an static year by year process.

the waves showing that approximately 5.5 – 6.6 % of individuals classified as legally disable.[19] Table 3 also shows that the percentage of individuals within the non-disability category who present some degree greater than 0 is almost negligent, with individuals who declare zero disability – we assume this to be individuals who have never been assess in a disability test – been the predominant majority. On the other hand, the distribution of the degree of disability within the group of those defined as legally disable shows a more even spread over the 30% to 89% mark.

Table 4 shows, also using the observations on a yearly basis, the interaction between one of the labour market outcomes considered – working versus not-working – and health status (disability versus non-disability). Clearly, due to relatively small number of disable in the sample (that, nevertheless, is the correct percentage relative to the true population), the percentage of working in the non-disable sample drives the estimated percentage in the full sample. Over the 18 years, there is almost no perceived change of the percentage of working in either of the two health status, with three quarters of the non-disable declaring to be working (employed, self-employed or in maternity leave) while the sample of disable shows a lower percentage of participation with some 40% of these declaring not to be working at the time of the survey (i.e., either registered unemployed, housework or early retirement). The final column in Table 4 shows that the distance in percentage between non-disable and disable is always significantly different than zero.

[19] See Note 1, Table 1 to see that this estimate is a very good approximation to the population projection.

***Table 4: Percentage estimates (standard errors) of working comparing samples according to disability status.***

| | | % OF WORKING SAMPLES FOR: | | | |
|---|---|---|---|---|---|
| **Year** | **Net Sample** | **Full Sample** | **Non-disable Sample** | **Disable Sample** | **One sided t-statistics.** |
| **1984** | **7074** | 68.0 (0.6) | 68.6 (0.6) | 59.3 (2.4) | 3.8 |
| **1985** | **6606** | 68.8 (0.6) | 69.4 (0.6) | 59.1 (2.5) | 4.2 |
| **1986** | **6404** | 68.8 (0.6) | 69.6 (0.6) | 55.8 (2.5) | 5.6 |
| **1987** | **6195** | 70.7 (0.6) | 71.2 (0.6) | 61.6 (2.6) | 3.9 |
| **1988** | **5843** | 71.6 (0.6) | 72.2 (0.6) | 60.5 (2.8) | 4.1 |
| **1989** | **5548** | 73.5 (0.6) | 74.3 (0.6) | 59.5 (2.8) | 5.3 |
| **1990** | **5460** | 73.4 (0.6) | 74.7 (0.6) | 54.4 (2.6) | 7.1 |
| **1991** | **5352** | 77.1 (0.6) | 78.1 (0.6) | 61.0 (2.9) | 5.8 |
| **1992** | **5219** | 77.3 (0.6) | 78.2 (0.6) | 61.0 (2.9) | 5.8 |
| **1993** | **5217** | 76.1 (0.6) | 77.4 (0.6) | 56.1 (2.9) | 7.2 |
| **1994** | **5083** | 75.2 (0.6) | 76.2 (0.6) | 58.3 (2.8) | 6.3 |
| **1995** | **5068** | 75.8 (0.6) | 77.0 (0.6) | 57.9 (2.8) | 6.7 |
| **1996** | **4971** | 75.8 (0.6) | 77.2 (0.6) | 56.0 (2.8) | 7.4 |
| **1997** | **4877** | 75.5 (0.6) | 76.9 (0.6) | 54.9 (2.9) | 7.5 |
| **1998** | **4684** | 75.9 (0.6) | 77.0 (0.6) | 58.0 (2.9) | 6.4 |
| **1999** | **4551** | 75.9 (0.6) | 76.9 (0.6) | 61.1 (3.0) | 5.2 |
| **2000** | **4326** | 76.0 (0.6) | 77.0 (0.7) | 59.8 (3.0) | 5.5 |
| **2001** | **4161** | 76.6 (0.7) | 77.7 (0.7) | 58.9 (3.1) | 5.9 |

Table 5 also shows – for selected years – the interaction between health status and the two labour market outcomes considered in the empirical section, namely annual labour income and per capita household disposable income. With respect to annual labour income, we consider two definitions; assigned labour income which includes zero income for those who declare a non-working status over the calendar year, and labour income for those who declare positive earnings. First, notice that for both non-disable and disable individuals, labour income and per capita household income has increased steadily over time. Accounting for zeros we observe that those in the non-disable sample are on average higher earners than their disable counterparts, although the difference is never significantly different than zero for any of the years considered; however, comparing sample medians, also allowing for zero incomes, shows that up to 1987 there was a higher probability of earning a higher income among disable individuals, relative to non-disable, while from 1990 onwards this probability reverses between the two health status groups. One result of this is that the earning of non-disable have increased, on average, 40% between 1984 and 2001, while the analogous increase for disable individuals is 21%. Once we account for positive earners only the roles between disable and non-disable seem to reveres, with disable showing slightly higher earnings than the non-disable population throughout all the years, except for the year 2001. This suggest that zero incomes deflate the income of disable substantially more than the effect these might have on the non-disable sample. In fact, averaging over the years considered, some 20% in the non-disable sample are counted as zero incomes, while the percentage for the disable with zero incomes is, on average, 35%. Nevertheless, the per year distance between disable and non-disable using positive income is never significantly different than zero. The final measure of income we consider is annual per capita net household income, thus taking into account all income receipts by all household members. We can think of it as a measure to compare the purchasing power of the two groups. Non-disable show higher levels of disposable income, but the difference between the two samples is relatively small and never higher than 3,000 Deutsch Marks per year for any of the years considered. Table 5 also shows a measure of relative inequality by estimating the inter-quartile relative to the median. With this measure we see that within sample – i.e., for either disable or non-disable – the distribution of labour income leads to less equality than the distribution of per capita disposable income: this is expected since the distribution of labour income might be a reflection of within sample productivity, while net household per capita disposable income reflects government intervention (thus correcting for inequality in the population). The interesting observation with respect to IQR/Q50 is that for labour income, and specially with respect to the sub-population of positive earners, there is more (productivity) inequality in the population of non-disable than in the disable counterpart. On the other hand, comparing between the two samples shows that the distribution of per capita disposable income leads to estimates between the two groups for IQR/Q50 that are never sufficiently different to suggest that the two groups differ in a significant way with respect to purchasing power inequality.[20]

Although both Table 4 and Table 5 shows the interaction between disability status and various labour market outcomes, neither tables provides evidence of a causal relation between disability and labour

[20] Table 5 makes reference to the sample selected according to SSC1. Table C5, Appendix C, shows similar estimates when the sample selection criteria refers to SSC2. Is worth noting at this point that for this selection, where everyone in the sample is a full time worker at the beginning of the three year sequence, the differentials in productivity inequality between samples and over time is much lower (as expected given the criteria behind SSC2).

market outcomes, but rather it summarizes evidence of possible differentials between disable and non-disable persons that need to be studied with more appropriate statistical tools.

***Table 5: Summary statistics for two definitions of labour income and per capita household income according to disability status (all estimates are in Deutsch Marks, base year 1999).***

| | Labour Income | | NON-ZERO Labour income | | Per Capita household disposable income | |
|---|---|---|---|---|---|---|
| **Year** | **Non-disable** | **Disable** | **Non-disable** | **Disable** | **Non-disable** | **Disable** |
| **1984** | | | | | | |
| **Mean (s.d)** | 27,300 (33,400) | 27,400 (27,200) | 38,700 (33,800) | 42,400 (22,500) | 28,600 (15,100) | 28,400 (17,000) |
| **Median** | 21,800 | 26,300 | 36,100 | 42,100 | 26,600 | 26,200 |
| **IQR/Q50** | 2.00 | 1.75 | 0.90 | 0.61 | 0.55 | 0.56 |
| **1987** | | | | | | |
| **Mean (s.d)** | 30,100 (32,500) | 30,000 (28,800) | 39,900 (31,800) | 44,700 (24,000) | 32,000 (16,600) | 30,500 (12,600) |
| **Median** | 25,700 | 28,600 | 37,100 | 45,600 | 29,800 | 29,700 |
| **IQR/Q50** | 1.81 | 1.80 | 0.96 | 0.61 | 0.53 | 0.52 |
| **1990** | | | | | | |
| **Mean (s.d)** | 35,100 (42,200) | 28,800 (30,100) | 44,100 (42,900) | 46,000 (25,700) | 35,300 (19,400) | 32,500 (14,100) |
| **Median** | 31,400 | 23,800 | 41,200 | 47,200 | 32,300 | 30,700 |
| **IQR/Q50** | 1.54 | 2.20 | 0.95 | 0.63 | 0.54 | 0.64 |
| **1993** | | | | | | |
| **Mean (s.d)** | 39,800 (39,600) | 31,600 (34,400) | 48,900 (38,400) | 51,100 (30,300) | 39,500 (20,200) | 37,400 (17,700) |
| **Median** | 37,700 | 24,500 | 46,100 | 48,200 | 36,700 | 35,400 |
| **IQR/Q50** | 1.40 | 2.27 | 0.91 | 0.72 | 0.58 | 0.60 |
| **1996** | | | | | | |
| **Mean (s.d)** | 40,900 (41,900) | 33,300 (35,100) | 50,200 (41,100) | 52,400 (30,600) | 39,400 (22,400) | 37,800 (17,900) |
| **Median** | 37,200 | 29,700 | 47,000 | 50,100 | 36,000 | 35,200 |
| **IQR/Q50** | 1.48 | 1.91 | 0.97 | 0.68 | 0.56 | 0.620 |
| **1999** | | | | | | |
| **Mean (s.d)** | 40,900 (40,000) | 35,700 (36,200) | 50,100 (38,700) | 51,800 (32,600) | 39,500 (21,600) | 37,800 (16,800) |
| **Median** | 37,200 | 31,500 | 46,800 | 51,000 | 36,200 | 35,200 |
| **IQR/Q50** | 1.51 | 1.89 | 0.98 | 0.68 | 0.57 | 0.63 |
| **2001** | | | | | | |
| **Mean (s.d)** | 43,800 (44,300) | 34,600 (35,000) | 52,500 (43,600) | 50,300 (31,500) | 41,600 (23,800) | 40,300 (17,000) |
| **Median** | 39,300 | 31,000 | 47,800 | 45,900 | 37,900 | 37,500 |
| **IQR/Q50** | 1.49 | 1.89 | 1.02 | 0.86 | 0.57 | 0.58 |

Note 1: IQR/Q50 refers to the relative Inter Quartile Range, weighted by the median. This estimate allows for a measure of inequality, while weighting by the median controls for overall shifts in the income distribution. If IQR/Q50=0, this reflects perfect equality, while IQR/Q50>0 reflects increased levels of inequality, with $IQR/Q50 \in [0,\infty)$. This measure is used with respect to per capita household income, since this constitutes a measure close to purchasing power inequality.

The sample of 10,995 forms the basis for the construction of the treatment and control (untreated) groups. Of those, only 1,642 are consistently observed between 1984 and 2001. Not all individuals contribute towards either control or treatment, as some might not comply with either group's definition, while other individuals might be observed either for less than three years or intermittently over the 18 years. Once we consider each sequential three years, the sample of interest becomes the number of individuals observed over the three years for each of the sequence, and these are distributed into three different groups: either controls, treated units, or non-contributing units (with respect to the control or treatment groups). Table 6 shows the sample size for each of the sequences $S_1$ to $S_{16}$, and how each of these samples is distributed between the three possible alternative groups.

***Table 6: Distribution of sequential observations between not –used individuals, untreated (control) group and treated group***

| Sequences | [t1, t2, t3] | Total number of sample points | Total number who do not contribute to either control or treatment | Total in the untreated sample, i.e., comparisons [AAA] | Total in the treated sample, i.e., treatments [ADD] |
|---|---|---|---|---|---|
| **S1** | **[1984, 1985, 1986]** | 5,573 | 471 | 5,056 | 46 |
| **S2** | **[1985, 1986, 1987]** | 5,329 | 402 | 4,927 | 0 |
| **S3** | **[1986, 1987, 1988]** | 5,183 | 397 | 4,737 | 49 |
| **S4** | **[1987, 1988, 1989]** | 4,923 | 409 | 4,482 | 32 |
| **S5** | **[1988, 1989, 1990]** | 4,928 | 629 | 4,267 | 32 |
| **S6** | **[1989, 1990, 1991]** | 4,700 | 384 | 4,257 | 59 |
| **S7** | **[1990, 1991, 1992]** | 4,595 | 379 | 4,214 | 0 |
| **S8** | **[1991, 1992, 1993]** | 4,598 | 334 | 4,225 | 39 |
| **S9** | **[1992, 1993, 1994]** | 4,427 | 356 | 4,071 | 0 |
| **S10** | **[1993, 1994, 1995]** | 4,472 | 378 | 4,040 | 54 |
| **S11** | **[1994, 1995, 1996]** | 4,401 | 408 | 3,965 | 28 |
| **S12** | **[1995, 1996, 1997]** | 4,323 | 386 | 3,905 | 32 |
| **S13** | **[1996, 1997, 1998]** | 4,159 | 375 | 3,760 | 24 |
| **S14** | **[1997, 1998, 1999]** | 4,022 | 327 | 3,666 | 29 |
| **S15** | **[1998, 1999, 2000]** | 3,850 | 322 | 3,504 | 24 |
| **S16** | **[1999, 2000, 2001]** | 3,703 | 304 | 3,378 | 21 |
| | **Totals** | | | 66,454 | 469 |

Note1: Sequences S2 (1985 to 1987), S7 (1990 to 1992) and S9 (1992 to 1994) are not used in the final estimates of Section 5 since they show zero counts of units with patter ADD over their corresponding 3 year periods.

As expected, the number of units in the control group far outweighs the number of treated units. Within each sequence, the situation cannot be improved given the existing panel. If we consider longer time periods per sequence (e.g., allow 4 instead of 3 years such that the outcome of interest is observed at $t_4$), the number of units drops for both control and treated units for each of the sequences, further reducing the number of observed units in the treated groups, thus becoming more of a problem.[21] On the other hand, considering sequences of two time periods only (so that our interest would fall on comparing labour market outcomes at $t_2$) would not allow sufficient time for disability policies to have an impact on the labour market outcome for newly diagnosed disable persons. The overall count in Table 3 shows a total of 469 treated units. As more waves become available this number could increase if we increase the number of sequences considered, as long as this provides a positive number of individuals with a patter

[21] See Footnote 16 and Appendix B.

ADD between non-disability and disability. Nevertheless 469 treated units appears to be a sufficiently large number of observations to make inference on the impact of such treatment on various definitions of labour market outcomes.[22]

Table 7 shows some comparative statistics for selected variables, comparing the average outcome of these at $t_3$ to the average observed at $t_1$, between controls (totalling 66,424) and treatments (totalling 469). So, for example, averaging over all sequences, 56.4% in the control group experience increases (in real terms) in annual labour income between $t_3's$ and $t_1's$, while 29.6% experience a decrease, with 14% suggesting no real change on annual earnings. Similar estimates for the treated group show that only 40.3% show increases in real earning between $t_3's$ and $t_1's$, while 42.2% experience a real decrease and 17.2% no change. Column 5 in Table 7 tests for any significant difference between percentages (increase, decrease, no change) between the two groups, for any of the variables considered. It is surprising that whichever measure of income we look at (either earnings or per capita disposable income), the percentage of individuals with increasing incomes is significantly larger for the control group, relative to the treated group, although the percentage difference between groups in terms of labour income is almost 4 times as large as the difference in percentage that see their per capita household income increasing. With respect to changing employment status the difference is also significant: 8.4% more individuals in the treated group change from employment to non-employment, relative to the control group where only 6.1% experience such a change. Likewise, 7.4% more units in the treatment group show no change from the status not-employed, relative to the control group, with 23.9% of the treated units remaining in a non-employment status between $t_3's$ and $t_1's$ while only 16.5% of the control units are classified as such. In

[22] As previously suggested (see Table 1, and Section 2 in general), the percentage of disables in each wave is similar to national figures on working disable of the population in West Germany. Because there is no statistical knowledge on the distribution of disable over three year periods, we need to assume that our data represents such distribution correctly, so that even with a small number of counts in the treated group over the three years, this is what we would obtain in the actual population. The only factor that would affect this assumption is attrition over time, and if attrition is not independent from disability. For example, if attrition is a sign of motivation, our

terms of satisfaction with work, 31.5% see their job satisfaction increase if inside the control group, relative to only 24.5% who declare increase job satisfaction in the treated group. Paradoxically, there is a significantly greater percentage of individuals in the treated group that have experience an increase satisfaction with their health (17.9%) compared to individuals in the control group (only 4%); however, the percentage that also experience a decrease in health satisfaction is larger in the treated group, relative to the control group, with 13.9% versus 4.7%, respectively.

***Table 7: Change over time (between $t_1$ and $t_3$, average over each of the two periods), for outcome variables (income, household income and employment), and two selected subjective measures (health and work satisfaction).***

| | | *Comparison Group [A(t1)A(t2)A(t3)]* | *Treatment Group [A(t1)D(t2)D(t3)]* | *One sided t-test* |
|---|---|---|---|---|
| **Annual labour income** | *increased* | 0.564 | 0.403 | 7.08 |
| | *stayed the same* | 0.140 | 0.175 | -6.96 |
| | *decreased* | 0.296 | 0.422 | -5.51 |
| **Annual per capita household net income** | *Increased* | 0.611 | 0.567 | 1.92 |
| | *stayed the same* | - | - | |
| | *decreased* | 0.389 | 0.433 | 1.92 |
| **Employment status (E= Employed) (NE=Not-Employed)** | *from E to NE* | 0.061 | 0.145 | -5.16 |
| | *stayed E over time* | 0.698 | 0.576 | 5.33 |
| | *stayed NE over time* | 0.165 | 0.239 | -3.75 |
| | *from NE to E* | 0.077 | 0.041 | 3.91 |
| **Average number of hours working per week** | *increased* | 0.337 | 0.215 | 6.40 |
| | *stayed the same* | 0.349 | 0.394 | -1.99 |
| | *decreased* | 0.314 | 0.390 | -3.36 |
| **Satisfaction with work (subjective)** | *increased* | 0.315 | 0.245 | 3.51 |
| | *stayed the same* | 0.352 | 0.371 | -0.85 |
| | *decreased* | 0.333 | 0.384 | -2.26 |
| **Satisfaction with health (subjective)** | *Increased* | 0.040 | 0.179 | -7.85 |
| | *stayed the same* | 0.912 | 0.682 | 10.68 |
| | *decreased* | 0.047 | 0.139 | -5.75 |

Note: Estimates for subjective satisfaction with work are based the restricted sample of those declaring to be working only.

To some extend, Table 7 provides some evidence on the impact of health status over time, for the various outcome variables as defined in Column1. If one could assume that disability is a shock that affects the population in some random fashion, then Table 7 would provide unbiased estimates on the effect of disability on variables such as the various incomes considered and employment participation. However, the assumption that disability affects individuals randomly is not plausible, thus Table 7 is not appropriate for policy analysis given that estimates do not control for factors that might have a simultaneous effect on

results can be biased towards the outcome of disables with higher levels of motivation that decide to remain in both health and outcomes. Instead, Table 7 gives some indication on the dynamic changes of the data and variables of interest.

An additional characteristics for the treated group is the distribution of the degree of disability that can take values from 0 to 29% in the first period, and from 30% to 89% in the second and third periods for each of the sequences. We think of disability as a characteristic that increases over time, from 0 to 100%, and does not have an effect on the control group since these are defined as individuals who show a non-disability status consistently over the three year period (for any sequence) while declaring zero degree of disability. On the other hand, those in the treated group can have a degree of disability between 0 and 29% in period $t_1$, and see this measure increase over $t_2$ and $t_3$, to be between 30% and 89%.[23] Table 8 shows the distribution of the degree of disability in terms of deciles over $t_1, t_2$ and $t_3$ accounting for all sequences with positive number of treated units.

***Table 8: Distribution of degree of disabilities***

| *Degree of disability* | *Distribution between degree brackets at $t_1$* | *Distribution between degree brackets at $t_2$* | *Distribution between degree brackets at $t_3$* |
|---|---|---|---|
| **0 degree** | 83.5% | - | - |
| **$0 <$ degree $\leq 10$** | 0 | - | - |
| **$10 \leq$ degree $\leq 19$** | 0.2% | - | - |
| **$20 \leq$ degree $\leq 29$** | 16.3% | - | - |
| **$30 \leq$ degree $\leq 49$** | - | 53.8% | 45.9% |
| **$50 \leq$ degree $\leq 69$** | - | 34.7% | 40.1% |
| **$70 \leq$ degree to 89** | - | 11.4% | 14.0% |

Note: Table is based on 469 treated. The percentages are the result of constructing the sample according to degree of disability, therefore, 0% of those in the treated group in the groups below 30% of disability occurs by construction.

Table 8, Column 1 shows that the vast majority of individuals who been non-disable become disable in subsequent periods are individuals who start up the process with no degree of disability. Although we have no information as to how the disability occurred, one possible cause to explain the status

the panel study.

[23] Recall that in our sample selection criteria one condition is that disability had to be between 0 and 89%, therefore beyond $t_1$ anyone in the treatment group has a degree of disability of at most 89%.

change is a random shock that the individual receives.[24] Only a very small number (around 78 out of 469 in the total sample of treated) declare to have some degree of disability greater than zero at $t_1$. At $t_2$, and by construction, anyone in the treated group has a degree of disability between 30% and 89%, with more than 50% belonging to the group where disability is between 30% and 49%. The difference in distribution of disability degree between $t_2$ and $t_3$ – Columns 3 and 4, respectively – shows a gradual increase in the diagnosed degree of disability for the individuals in the treated group, with categories of degree beyond 49% increasing in percentage over the category with degree between 30%-49%.

### *3.2 Control variables*

The GSOEP100 provides rich quality data at the individual level. The panel contains information on key variables that reflect the characteristics that drive the chances of individuals on becoming disable while having an effect on individual's labour market outcomes. These variables are needed because it is conditioning on them that allows for the assumption of independence between labour market outcomes in the control group and disability treatment.[25] The selection of such variables could be done following some international guidance on the classification of causes or underlying conditions on disability, for example "*The WHO international Classification of Disease and Related Health Problems (ICD-10)*". Surveys and census in countries that have used such guidelines[26] will often include a question – very similar in all of

---

[24] This been the case our analysis would be more valid still because then it is truly the case that pre-disability characteristics are independent from outcomes, and that once controlling for all observed characteristics, control and treated are otherwise identical. The distribution between degrees of disability for the SSC2 sample is even more inclined towards the zero-disability group in the first period (see Appendix C). On the other hand, it could be that the person is disable in some progressive way (for example, impaired hearing), but has never voluntarily submitted to a disability test that might have given that person some degree of disability between 0 and 29%. Once they proceed to a test after the $t_1$ period, they jump from 0 to, say, 50%. This however, is not a random shock as would be loosing a limp, for example, due to a sport or work injury.

[25] See Section 4.

[26] For example, the 1993 Australian Survey in Disability, the 1996 Household disability survey in New Zealand, or the 1998 Netherlands Health Interview Survey, but to mention a few.

them – where individuals are able to classify their impairment/limitations among a set of categories.[27] Based on examination of health surveys in various countries over time, the United Nations Statistical Division[28] has proposed a short list for classifying causes of disablement which includes three categories relating to genetics and acquire diseases (infectious and parasitic diseases, congenital anomalies and prenatal conditions, other diseases related conditions), four categories with references to external injuries (motor vehicle accidents, other transport accidents, accidental poisoning, injuries from activities – falls, fires and wars), and a category which includes all disability causes related to environmental factors.

We follow the UN guidelines in selecting those variables in the GSOEP survey that might have an effect (a priory) on the probability of people becoming disable. The selection of variables can be classified into five categories, namely $(i)$ Traffic, $(ii)$ Genetic, $(iii)$ Labour market classification, $(iv)$ Leisure activities and $(v)$ Demographic and social-economic status (SES). In terms of 'Traffic', we think that the degree of urbanisation is positively correlated to the chances of becoming disable, while urbanisation is also an important determinant for the local employment rate and labour income of any given area in a country. By 'genetics' we mean any endowment which parents can pass on to their children that might affect both the chances of the child to become disable as an adult, but also their work status. Variables such as the level of parental education would enter this category: one would assume that, on average, parents with more education are better at transferring information on safety to their children (e.g., using seatbelts when driving) that will affect the probability of the child on becoming disable once the child becomes an adult. Another example is that parents with more education might be better at processing information, including the importance of nutritional needs of growing children, so that parental education may have a direct impact on the child's capacity to avoid illness in adulthood which are associated with

[27] The question differs among countries with respect to the detail given to each category. For example, in the 1998 Netherlands survey, individuals are asked to classify disability via illness between congenital or occurring at birth, illness of childhood, or illness of old age; the 1996 New Zealand survey is similar but provides further categories by age groups, distinguishing with reference to illness due to either psychological or physical abuse.

[28] "Guidance and Principles for the development of disability Statistics"(2001), UN Development of Economics and Social Affairs, UN Publications, Statistical Division, http://www.un.org/depts.unsd.

poor environmental growing-up conditions. At the same time, parental education has a direct impact on the child's ability and education level, therefore directly affecting the child's work status once they reach adulthood. Ideally the category 'Genetics' would also include objective measures of parental health status (e.g., the parents disability status, chronic illness, etc.), but such variables are not found in the GSOEP survey, thus a measure of the highest level of education achieved by mother and father are the only two variables that we include in this category. The category 'Labour market status' includes variables such as occupational category (e.g., blue collar worker), as well as variables specifically related to work and work activities (e.g., number of hours work, level of risk of work activities, time in the firm and size of firm's workforce). The category 'Leisure' refers to activities outside work hours. This category includes information such as participation in sports, since we can think that such activities are often associated with the status of individuals (i.e., income, time availability), while the risk element in leisure activities can also affect disability status. Finally, demographic and socio-economic (SES) related variables (e.g., family size or number of dependent children, age, education, marital status, etc.), are also important controls that can affect both the treatment and outcomes to be analysed in the empirical section. Table 9 lists all variables used in the estimation process (see Sections 4 to 6) according to categories, while Appendix D (see Footnote 12) provides a more detail description of these variables and the construction of any secondary variable derived from them.

Table 9 shows that the GSOEP survey is excellent providing information on labour market status and activities, demographics and SES. As shown in Appendix A, the survey is also informative with respect to objective information on disability so that we can identify legally disable individuals without uncertainty.[29] One problem with respect to the available information in the survey is the lack of information on individual's health habits (e.g., smoking habits, alcohol consumption or diet), since information for this variables is only available intermittently over the last three year of the survey. Although this health habits

[29] This improves on many studies of disability where the degree of disability is often not well defined and subjectively interpreted by either researcher, data collection and subjective data presentation techniques, or surveyed individuals who are asked to declare subjective satisfaction with health that might not be easily interpretable or comparative between surveyed units.

might certainly have an impact on health, as well as been related to labour market status, due to the way we construct the treatment and control groups, the information in the panel at this point is insufficient to enter our analysis. We do, however, have a variable on sports practice of individuals that can be thought to be of some proxy to underlying information on the individual's health habits.

***Table 9: Classification of covariates***

| Leisure | Traffic | WORK related variables | Socio-economic characteristics | Genetics (Parental background) |
|---|---|---|---|---|
| (1) Sports<br>(2) Motivation | (1) Land (dummies for 'Berlin', 'other cities', 'Non-city area')<br>(2) Degree of urbanisation (various dummies to classify according to density of inhabitants, from very low if number below 1000 per area, to very high if number of inhabitants is greater than 100,000 per given area) | (1)Satisfaction with work<br>(2) Blue collar<br>(3) White collar<br>(4) Civil servant<br>(5) Self-employed<br>(6) Farmer<br>(7) Level of risk at work (jointly use of collar and ISCO code to identify primary, secondary and tertiary industry, and risk of activity performed within industry)<br>(8) Size of company (number of workforce)<br>(9) Time since individual started working in present company<br>(10) Average number of Hours spent at work per week | (1) Age<br>(2) Gender<br>(3) Household size<br>(4) Number of kids<br>(5) Partner indicator<br>(6) Years of education<br>(7) Status of house ownership. | (1) Father's education<br>(2) Mother's education<br>(3) World origin (region of the world where person was initially brought up) |

# 4 Parameters of interest and their identification

The question we aim to answer is "What is the effect of becoming disable, for those who become disable, on their labour market outcomes, compared to the hypothetical state of not having received the impact of disability?" This question targets the causal relation from disability to outcome (i.e., labour earnings, participation), and can be answered using Rubin (1984) potential outcome approach to causality.

The population of interest in each $S_j$ sequence, for $j = 1,...,J$ , is defined over periods of three years $(t_1, t_2, t_3)$ and is represented by a sample of size $n_j$ . Individuals $i \in n_j$ only enter the sample if they are observed over the three years and, at the same time, comply with our sample selection criteria (see Section 3). In our case $J = 16$ . For each unit we have health status information (either disable or non-disable) at each of the three years for any of the sequences. The dynamic (un-)change between non-disability and disability identifies three possible groups within each $n_j$ : (1) individuals are controls units when their health patters is $A_{t1}A_{t2}A_{t3}$ over the sequence $S_j$ or, (2) individuals are treated units, so that the patter in this case is $A_{t1}D_{t2}D_{t3}$ or (3) individual's health pattern does not comply with neither of these two alternatives[30]. Whereas only the treated and the control units are used when estimating the actual effect of disability on labour market outcome for the disable, the underlying population is represented by all three groups in $n_j$ and, therefore, we need to use all three groups when estimating uncertainty due to sampling error (see Section 4.1).

Let $T_{i,j}$ be a binary assignment indicator that determines whether unit $i \in n_j$ gets the treatment ($T_{i,j} = d;\ \Rightarrow A_{t1}D_{t2}D_{t3}\ over\ S_j$) or not ($T_{i,j} = a;\ \Rightarrow A_{t1}A_{t2}A_{t3}\ over\ S_j$). Omitting the suffix $j$ for simplicity, let $Y_i^d$ and $Y_i^a$ be the potential labour market outcomes associated with the treated and untreated (control) states, respectively. The notion "potential" is used to emphasis that only one of $(Y_i^d, Y_i^a)$ is observed for every unit in the sample. Each unit that is either in the control or treatment group is an individual $i \in n$ identified as non-disable in period 1 $(t_1)$. The sub-sample of treated units is constructed so that an individual observed to be disable at $t_2$ is also observed as disable at $t_3$: $Y_i^d$ is the actual (observed) outcome $(Y_i)$ at $t_3$ associated with an individual $i \in n$ with such health pattern. Likewise, the actual outcome $(Y_i)$ at $t_3$ is $Y_i^a$ for a unit observed to be non-disable at $t_2$ who, also by construction, is observed to be non-disable at $t_3$.

---

30 For example, an individual is AAD over a sequence $S_j$ cannot count as either control or treatment over this three year period. If anything, this person might show a pattern AADD if we look one year ahead, so that the person would count as a treatment unit at $S_{j+1}$. Other combination – always over three-year periods – of disability status within $n_j$ not contributing towards either control or treatments are ADA – a combination that is assumed to be error coding since disability status should be irreversible –, as well as any combination that shows a status D at $t_1$, since by definition all those who eventually count as control or treatment are non-disable in each of the first period of each three year sequence. Thus, combinations DDD, DAD, DAA or DDA over any given sequence belong to the third group who are neither control or treatments).

Our parameter of interest, $\vartheta^0$, is the mean effect (at $t_3$) of receiving the impact of a disability shock, rather than not receiving the shock, on those individuals who having become disable at $t_2$ do in fact receive the impact of such an status thereafter (e.g., receiving the impact of the policies aimed at the disable, modification of behaviour with respect to labour market activities, etc.). This parameter is known in the literature as the "Average Treatment Effect on the Treated"(ATET), and can be expressed as:

$$\vartheta^0 = E[Y^d - Y^a \mid T = d] = E[Y^d \mid T = d] - E[Y^a \mid T = d] \qquad (1)$$

Clearly, $\vartheta^0$ is not identified by the data, since identification of the causal effect would require the observations of the counterfactual outcome to $Y_i^d$ (i.e., $Y_i^a$) for each $i$ unit in the treated sub-sample, thus allowing us to estimate $E[Y^a \mid T = d]$. Assuming that the probability of becoming disable is a random process, i.e., $E[Y^a \mid T = a] = E[Y^a \mid T = d]$, would solve the problem since the average outcome of the control sub-sample could be used as the (average) counterfactual outcome for the treated units. However in light of the evidence discussed in Sections 1 and 2, the random selection assumption is certainly violated in our context, specially with characteristics such as occupational sector clearly differing with respect to incidence of disability, since these characteristics are key determinants of labour market outcomes (see Table 1, Section 2).

Section 3 suggests that our data was very informative with respect to observed characteristics that might determine both health status and outcomes of interest (i.e., employment status, earning and income). Assume that a set of characteristics given by the vector $X$ is both sufficiently informative and unaffected by the treatment itself. Then, identification of $\vartheta^0$ is possible since conditioning on $X$ implies that within sub-groups (as defined by $X$), being a control (or not) is unrelated to what the outcome would have been if you had become disable (or not). This assumption is known as the Conditional Independence Assumption (CIA) and is formally given by:

$$(Y^d, Y^a) \perp T \mid X = x; \qquad \forall x \in \chi; \quad \chi \subseteq \mathbb{R}^p \qquad (2)$$

Therefore, $E[Y^a | T = d, X] = E[Y^a | T = a, X]$ and $\vartheta^0$ is identified such that,

$$\vartheta^0 = E[Y^d | T = d] - \underset{X|S=d}{E}[E[Y^a | T = a, X = x] | T = d] \qquad (3)$$

The CIA is a workable assumption as long as it holds for the available $X$ set, but does not account for unobserved characteristics that may also play a role in selection. As previously suggested, one further condition for the CIA is that all characteristics in the set $X$ have to be unaffected by the treatment itself; a violation would lead to endogeneity between control and outcome variables. For example, an individual who becomes disable at $t_2$ may decide to engage in further education, a decision that might affect her labour market outcome at $t_3$. In this case, if we allowed for years of education as measured at $t_3$ to enter the conditioning set $X$ in (3) we would end up with endogeneity problems because the treatment, which affects the outcome, determines the controls: in fact, once a person becomes disable, one possible disability policy is to engage these individuals into programs that might increase the number of years in education, so to become more marketable with respect to labour market activities. Clearly, if we want to estimate the effect of disability policies on their labour market outcomes for the disable, we need to avoid controlling for observed characteristics that might be the effect of the policies themselves. To make sure this is the case we need to use a set of $X$ which is not influenced by the treatment. In our case, this is already given when we construct our treatment and control units, i.e., by construction all individuals (treated and controls) are non-disable in period $t_1$, and therefore, conditioning on $X$ at $t_1$ (i.e., conditioning on $X_1$) implies that this exogeneity condition is fulfilled. With this, a more complete version of (3) is given by:

$$\vartheta^0 = E[Y_3^d | T = d] - \underset{X_1|S=d}{E}[E[Y_3^a | T = a, X_1 = x_1] | T = d] \qquad (4)$$

where (4) implies that we are interested on comparing outcomes at $t_3$ between treatment and comparison units, given that they shared similar characteristics a the point where both control and treatment units where in one single state of the world (non-disability).

$\vartheta^0$ can be estimated using the sample analogue, provided that for every treated unit there is a comparison unit in the control sub-sample with similar $X_1$ characteristics. This is known as the common support condition, which for ATET in our particular application is defined as $P(T = d \mid X_1 = x_1) < 1$; $\forall x_1 \in \chi_1 \subset \mathbb{R}^p$. The implication is that there is a common overlap between the distribution of the set $X_1$ in the two states.

When $X_1$ is of high dimension, estimation of $E[Y^a \mid S = a, X_1]$ using distribution free techniques such as Kernel based nonparametric methods is subject to the so-called curse of dimensionality (i.e., very low density per cell) and, therefore, increases imprecision in the parameter estimates, specially in the tails of the distribution. However, Rosenbaum and Rubin (1983) show that it is not necessary to compare observations with the same value of $X_1$, but it is sufficient to compare observations having the same conditional treatment probability, $P(T = d \mid X_1 = x_1) = p(x_1)$, where $p(x_1)$ is also known as "the propensity score". Conditioning on $p(x_1)$ rather than $X_1$ itself reduces the estimation problem to a one dimension so that the estimate of $\vartheta^0$ will be based on:

$$\vartheta^0 = E[Y_3^d \mid T = d] - \underset{p(x_1)|S=d}{E}[E[Y_3^a \mid T = a, p(X_1) = p(x_1)] \mid T = d] \tag{5}$$

### 4.1 Matching methods

Matching on the propensity score can be done in different ways. The key difference of these methods is the weight assigned to each observation in the control (or comparison) group. All matching methods are based on the following form:

$$E_n[Y^a \mid P_n(X_i)] = \sum_{c=1}^{C} w\left(p_{ni}(x_i), p_{nc}(x_c)\right) Y_c^a \tag{6}$$

where $c = 1,\ldots,C$ is the index for the control group and $i = 1,\ldots,I$ is the index for the treatment group, with $n$ referring to the sample size (of any give sequence). The expectation in (6) is taken over all $c \in C$

control individuals for each $i^{th}$ individual in the treatment group, therefore the counterfactual outcome for each treated unit is a weighted average of the outcome of the untreated group, where $w(.)$ is the interpretation of such weight. Different weighting methods (i.e., matching methods) imply different ways to weight the potential counterfactual observations, but also different ways to account for the common support problem.

The empirical section uses two alternative matching methods on the propensity score, leading to two comparative estimates of the average treatment effect on the treated for each of the three labour market outcomes of interest (working status, labour income and per capita household disposable income). The first matching method consists on comparing the estimate of the propensity score of each unit in the treated sample to the estimated propensity score of all control units, and assigning a weight of one to the observation in the control group whose estimated propensity score is closest to a treated unit, for each treated unit. This method is known as Nearest Neighbour matching, with $w(.)$ in (6) expressed as:

$$w\left(p_n(x_i), p_n(x_c)\right) = \begin{cases} 1 & if \quad c = \arg\min_{k \in C}\{|\, p_n(x_i) - p_n(x_k)\,|\} \\ 0 & otherwise \end{cases} \tag{7}$$

For cases showing a large overlap between the distributions of the two estimated propensity scores alongside a control pool much larger than the treated pool, it is common practice to follow the statistical literature and match without replacement; each observation in the control group is used at most once. Otherwise not using observations again might lead to the violation of the CIA, thus matching with replacement might be a more appropriate practice (see Black and Smith (2002) or Gerfin and Lechner (2002) for example). Weighting using the matching technique as given in (7) leads to different estimates according to how the absolute distance between estimated propensities accounts for the common support problem. As it stands, estimates based on (7) do not specify what constitutes an appropriate distance before an observation in the control group becomes a counterfactual for an observation in the treatment group and, therefore, (7) does not take the common support problem into account. To give an example, it might be that the $i^{th}$ observation in the treatment group shows a distance of 0.90 (with respect to estimated propensity scores) with the $k^{th}$ observation in the control group, where the propensity score is always $\in [0,1)$. If such distance is what minimises the argument in (7) for the $i^{th}$ treatment observation with respect to all control observations, then (7) will match the outcome of interest from the $k^{th}$ control as counterfactual for the $i^{th}$ treatment unit, even if by such distance of 0.90 the indication is that the $k^{th}$ control is a bad match for the $i^{th}$ observation in the treated sub-sample. To solve this problem, an alternative based on (7) is to match such that $c = \arg\min_{k\in C}\{| p_n(x_i) - p_n(x_k) | \leq \kappa\}$, where $\kappa$ is defined as a "calliper", thus accounting for the common support problem given that the calliper imposes a condition on the overlap between the density estimates of the propensity scores. In our case using a calliper or not has almost negligent effects on the matching process. For example, with $\kappa = 0.10$ only 3 treated units fail to find a match in the control group, whereas with $k = 0.05$ the number of un-matched treated units increases to 25. In both cases the effect of dropping these non-matched treated units on the final estimate of the ATET is almost negligent and, therefore, when using the Nearest Neighbour matching estimator we base our results on (7) without calliper.[31]

The second matching method considered in the empirical section is Kernel matching, with the analogue to (7) given as

$$w\big(p_n(x_i), p_n(x_c)\big) = \frac{K\left[\dfrac{p_n(x_i) - p_n(x_c)}{h_n}\right]}{\sum_{k=1}^{C}\left[\dfrac{p_n(x_i) - p_n(x_k)}{h_n}\right]} \tag{8}$$

[31] This results because there is an almost perfect overlap between estimates of the propensity scores for control and treated for each of the periods 13 periods considered, so that in our particular application we have no common support problem. See Appendix F for a graphic comparison between densities of propensity score estimates of controls and treated sub-samples, where such estimates are based on SSC1. The same nesting between propensity scores would be observed if we use those based on the SSC2 selection criteria: this follows since SSC2 is a sub-sample from SSC1.

where $K[.]$ is the kernel function and $h_n$ stands for the bandwidth. Relative to the Nearest Neighbour, a Kernel function might assign a non-zero weight to more than one observation (if not to all observations). As with the Nearest Neighbour, each observation in the control group is weighted according to distance between estimated propensity scores, but whereas with the Nearest Neighbour all control observations are simultaneously compared to each unit in the treatment group, in this case of using a Kernel the weights are determined by the distance within a sub-group of control observations where such sub-groups are determined by the bandwidths. In practice we choose the bandwidth small enough with respect to the variation in the density of the propensity score estimate for the control sample. The usual choices of bandwidths are based on either a normal approximation (see Silverman, 1986) or estimating a bandwidth by Cross Validation (see Haedler and Marron, 1985). We have used each of these two bandwidths on two types of Kernel function, namely, a Gaussian Kernel and Epanechnikov Kernel. Whereas with Gaussian Kernels there is no restriction on the support (for the weighted distance between treated and control units) the Epanechnikov Kernel "trims" away treated units at the tail of the density, thus accounting for the common support problem. Given that for our particular analysis there is no common support problem, (see Footnote 32), comparing matching estimates between Gaussian and Epanechnikov Kernels resulted negligent difference that have no effect on our final policy recommendation. Thus, our result section show only estimates using Gaussian Kernels with bandwidths estimates based on Silverman's (1986) normal approximation.[32]

[32] Appendix H shows estimates of the Bandwidth using the normal approximation, and compares these to estimates using a cross-validation method, for each of the time sequences under study. Given the almost negligent difference between these two choices of bandwidth it is no surprising that results of estimating Gaussian Kernel do not differ significantly according to which of the two bandwidth is used. See Black and Smith (2002) for a discussion on the relative importance of choosing either Bandwidth or Kernel functions in the weighting process.

No particular matching method (i.e., Nearest Neighbour of Kernel weighting) can be thought to be best. However, weighting with a Kernel function implies that more than one (if not all) observations in the control sub-sample might be given a non-zero weight, relative to the Nearest Neighbour where the full weight fall in one control unit. This leads to a difference between the two methods when computing standard errors, since the use of more observations in the case of a Kernel function reduces the variance in estimation, but always at the expense of inducing a bias due to the combined weighting. In the results section this theoretical point is visible in that standard errors on estimates of the ATET when matching by the Nearest Neighbour are always larger than when matching with a Gaussian Kernel. In practice, the precision of the estimated parameters is obtained by means of a naïve bootstrap procedure that consists on re-sampling with replacement from the original sample that we assume to be a random draw from the population. Notice that estimates of the ATET are based on the two sub-samples of treated and control units only. However, the full sample for each sequence, $n_j$ for $j = 1,...,16$, is defined by observations over a three year period, including those whose health pattern over time means that they are neither control or treated units. In our bootstrap procedure, the empirical distribution is obtained by sampling with replacement from the full sample $n_j$ for each bootstrap draw, given that all units inside $n_j$ are assumed to be equally likely to be drawn in the underlying population. This process reduced the probability of optimal matching for each treated unit for any given bootstrap draw, although in the limit (i.e., with a sufficiently large number of draws) the empirical distribution of the ATET with a naïve bootstrap process should mimic the underlying distribution in the population. We obtain the empirical distribution of the error term by re-sampling (with replacement) 500 times from $n_j$, for each $j = 1,...,16$.[33]

[33] See Appendix G for an illustration of the density for these empirical estimates.

# 5 Results

In this section we present estimates of the impact of becoming disable on three different labour market outcomes: working status (versus not working), annual labour earnings and per capita household disposable income. These estimates are obtained by matching on the propensity score.

Because we have a binary set up (two possible states, $ADD$ versus $AAA$), the propensity score $P(T = d \mid X_1)$ is estimated using a Probit model, where $T = d$ indicates $ADD = 1$ – i.e., a non-disable individual at $t_1$ becomes disable at $t_2$ and is further observed as disable at $t_3$. The conditional set is indexed with suffix 1 to indicate that we condition on pre-disability variables, thus solving the question of endogeneity between state dependent and explanatory variables (see Section 4 for a detail account). Section 3 showed that we have 16 different sequences of three years each, with a set of control and treatment units for each of them. We treat our data as having 16 independent sets of information.[34] Therefore we estimate the propensity score for both treatment and controls by estimating a Probit model on each of the valid $j$ sequences [35], and use such estimates to apply matching methods independently to each sequence.[36] The specification of our Probit model is based on all variables shown in Table 9, allowing for square terms and interactions. As an example Appendix E – complemented by Appendix D – shows

---

[34] Notice that we are trying to match each individual treated unit to a counterfactual that will be the best match in the group of control units. We can think that one conditional variable for the matching process is that the treated and counterfactual have shared the same macroeconomic conditions while having had access to similar government policies that affect the workings of labour markets. By comparing treated to control units within sequences (as opposed to allowing all controls in any $j$ sequence to be comparison units in other $g = 1,...,16 : g \neq j$ sequence), we are already braking the sample of controls and treatment units into cells with similar characteristics, in the same way as controlling – for example – for educational characteristics, would group controls and treatment units with similar number of years into education. What we do is to stratify the full sample according to characteristics as in Mueser et al. (2003), and our stratification takes into account sequential time variations.

[35] Table 6 – Section 3 – shows that sequences S2, S7 and S9 end up with zero number of treated units. This sequences are not valid for the final estimates, so that even if j=1,…,16, the valid number of sequences is 13.

[36] See Appendix B for a detail account of the algorithm leading to the propensity score estimate and the choice of counterfactual for all treated units over time.

the result of estimating the model for the first sequence (1984 to 1986). Score tests and goodness of fit test are used to find the appropriate variable specification for each of the estimated Probit in each of the samples as defined by three year sequences. For any sequence the choice of variables is always based on the unrestricted model, as is the case for the first sequence, the results of which are shown in Table E1, Appendix E.

Using the estimated Probit coefficient for each of the sequence, the propensity score estimates are based on $\hat{P}(T=d \mid X_{1,i}) = \Phi(x_{1,i}\hat{\beta}_s)$ and $\hat{P}(T=d \mid X_{1,c}) = \Phi(x_{1,c}\hat{\beta}_s)$ for treatment and control groups, respectively, where the suffix $S$ indicates sequence. Appendix F shows a figure for each of the time sequences that enters the estimation process, comparing for each of them the paired estimate of propensity scores.[37] All figures – except Figure 11 – show that the two densities overlap, therefore, for our particular selection of covariates, the common support is well defined for the treated group with respect to the comparison group for all sequence sets. Figure 11 might be a consequence of the small sample size of the treated group in this case.

We estimate and compare ATET using two alternative matching methods, namely, Nearest Neighbour without calliper and Gaussian Kernel with a normal approximation bandwidth. Using the SSC1 sample, the range of value (over all estimable sequences) for the bandwidth by normal approximation is between 0.001 and 0.006, with what seems to be an outlier of 0.05 for the last sequence of time. Similar values apply to the SSC2 sample (see Appendix H for a detailed explanation on the bandwidth choice and estimates of the bandwidth for each $S_j$).

For each of the matching methods, we estimate the impact of disability on the population of the disable, for each of the three different labour market outcomes considered. Table 12 shows the results of these estimates for the sample SSC1, whereas Table 13 shows similar results applied to the sample SSC2

---

[37] Appendix F refers to the population represented by SSC1. Analogous estimates for the population represented by the SSC2 sample are available from the authors on request: they suggest similar conclusions than those under SSC1, although in the case of SSC2 estimates of the propensity score for the 11$^{th}$ sequence imply good matching on the propensity score.

(i.e., a sub-sample of SSC1 working full time at $t_1$ of each $S_j$). We comment on each of these two tables separately and then contrast both sets of results.

Table 12 makes inference on the population of prime age individuals. For the outcome "working" (versus "not working") we see that 61.6% who are treated (by a disability impact) over a three year period are actively working at the time of the survey, whereas their counterfactuals in the non-treated population show a percentage of actively working between 64% and 71.3%, a range that is due to difference in matching methods. The result of these estimates suggest that becoming disable can lead to a significantly lower probability of been in employment, with as much as 9.6% differential. In terms of the outcome annual labour income, the disable population shows an average of DM32,300 when measured two years after the disability kicks in. Their counterfactuals in the population show average year earnings that range between DM 35,250 and DM 38,230. This means that once a person becomes disable their potential for productivity earnings is significantly reduced, to the point that a non-disable person can earn as much as 16% more than a disable person (who is otherwise identical in observed characteristics). Furthermore, once we take sampling variation into account our estimates suggest a differential that is statistically significant at least at a 0.10 critical level. The final outcome measure of interest is net per capita household income. Since this is a measure that reflects household composition and government intervention, we think that comparison of treated and control corresponds to a relative measure of purchasing inequality (or well-being): this is also reflected since in this case, the distance between treated and counterfactuals is not as large in magnitude as was in the case of earnings, with a differential that ranges between DM 980 and DM 1,630, with counterfactuals consistently showing to have larger per capita disposable income for any of the matching methods employed. Clearly, the fact that there are significant differentials between non-disable and disable for all three outcome measures, and the fact that such differentials are statistically significant, are evidence which do not support the correct functioning of disability policies, the aim of which is precisely to reduce both employment and wage differentials between the two sub-sets of individuals in the population. In order to further investigate if this conclusion is robust to working status, we perform an identical analysis but allowing only for individuals who, on a sequence by sequence basis, declare to be full time employed at $t_1$ (the pre-disability stage). We think of these individuals as people who are more aware of working conditions, including policies that aim at the disable, and who might be more incline to take advantage of such policies in the event of receiving a disability shock. For this reason one would think that with this population the disability impact on their labour market outcomes might be lower (relative to impact experience by all those in SSC1). Table 13 shows the result of estimation for SSC2. In terms of the variable "working" versus not working we see that, on average, after a three year period, those who become disable are more likely to be employed than in the case where we accounted for the full prime age population (i.e., around 81% versus the 61.6% previously observed in Table 12). The effect of disability on employment for SSC2 is only 0.4% lower than that for the SSC1 selection, with an estimated impact showing that disable can experience a 9.2% reduction in the probability of been employment, relative to their counterfactuals in the population. However, in the case of SSC2 the estimate impact of disability on the chance of working has a more precise range of values, moving between 8.5% and 9.2%, whereas for SSC1 the impact ranged between 2.3% and 9.6%: this reflects a more homogenous employment status between the SSC2's counterfactual selection, relative to the similarity in employment status of controls in SSC1 at $t_3$, which can in turn be taken as evidence that the bias induced by the choice of Kernel as matching method is perhaps less evident in SSC2 estimates than in SSC1. In all cases, the impact of disability on employment is adverse to the disable persons as well as been statistically significant.

Similar comments will apply to the outcome labour income. In the case where only full time workers at $t_1$ are considered (SSC2), both groups of treated and counterfactual show larger yearly earnings, on average, than in the SSC1 sample. However, the earnings differentials between treated and non-treated is now increased considerably, since this can be as high as 20% between disable and non-disable (who are thought to be otherwise identical). Finally, comparing averages of per capita household income between SSC1 and SSC2 shows that at least government intervention leads to no differential in disposable income between those who are full time and the rest of the prime age population: for example, treated units in SSC1 show an average per capita net income of DM 34,900 whereas for the SSC2 selection this is estimated at DM 36,660. However, comparing the treated (disable) group to their counterfactual within

SSC2 shows a significant difference between the two groups, with a differential that can be as large as 6%. Again, as with all other outcomes, the distance is statistically significant for most matching methods employed, at least at the 10% level of significance.

***Table 12: Estimates of Impact of disability on labour market outcomes (ATET, SSC1)***

| | **OUTCOME:** **ACTIVE WORKING STATUS, expressed in 100% base (standard errors in brackets)** | | | |
|---|---|---|---|---|
| | **$E[Y^{ADD}\|S=d]$** | **$E[Y^{AAA}\|S=d]$** | **$\hat{\vartheta}$ (impact estimate)** | **95% Confidence Interval on $\hat{\vartheta}$** |
| **NEAREST NEIGHBOUR with no calliper** | 61.6 (2.2) | 64.0 (5.1) | **-2.3 (5.0)** | [-17.0 : 2.8] |
| **GAUSSIAN KERNEL with bandwidth by normal approximation** | 61.6 (2.2) | 71.3 (2.3) | ***-9.6 (2.1)*** | [-13.1 : -4.9] |
| | **OUTCOME:** **ANNUAL LABOUR INCOME (in Deutsch Marks, Base 1999)** | | | |
| | **$E[Y^{ADD}\|S=d]$** | **$E[Y^{AAA}\|S=d]$** | **$\hat{\vartheta}$ (impact estimate)** | **95% Confidence Interval on $\hat{\vartheta}$** |
| **NEAREST NEIGHBOUR with no calliper** | 32,300 (1,330) | 35,250 (3,380) | **-2,960 (3,330)** | [-11,502 : 1,550] |
| **GAUSSIAN KERNEL with bandwidth by normal approximation** | 32,300 (1,330) | 38,230 (1,420) | ***-5,940 (1,320)*** | [-8,470 : -3,390] |
| | **OUTCOME:** **PER CAPITAL NET HOUSEHOLD INCOME (in Deutsch Marks, Base 1999)** | | | |
| | **$E[Y^{ADD}\|S=d]$** | **$E[Y^{AAA}\|S=d]$** | **$\hat{\vartheta}$ (impact estimate)** | **95% Confidence Interval on $\hat{\vartheta}$** |
| **NEAREST NEIGHBOUR with no calliper** | 34,900 (660) | 35,920 (1,480) | **-1,000 (1,590)** | [-4,970 : 1,295] |
| **GAUSSIAN KERNEL with bandwidth by normal approximation** | 34,900 (660) | 36,450 (490) | ***-1,530 (660)*** | [-2,980 : -320] |

*Table 13: Estimates of Impact of disability on labour market outcomes (ATET, SSC2)*

| | OUTCOME: ACTIVE WORKING STATUS, expressed in 100% base (standard errors in brackets) | | | |
|---|---|---|---|---|
| | $E[Y^{ADD}\|S=d]$ | $E[Y^{AAA}\|S=d]$ | $\hat{\vartheta}$ (impact estimate) | 95% Confidence Interval on $\hat{\vartheta}$ |
| **NEAREST NEIGHBOUR with no calliper** | 80.9 (2.3) | 89.4 (2.7) | **-8.5 (3.4)** | [-15.7 : -2.4] |
| **GAUSSIAN KERNEL with bandwidth by normal approximation** | 80.9 (2.3) | 90.0 (1.4) | ***-9.2 (2.6)*** | [-15.2 : -4.6] |
| | **OUTCOME: ANNUAL LABOUR INCOME (in Deutsch Marks, Base 1999)** | | | |
| | $E[Y^{ADD}\|S=d]$ | $E[Y^{AAA}\|S=d]$ | $\hat{\vartheta}$ (impact estimate) | 95% Confidence Interval on $\hat{\vartheta}$ |
| **NEAREST NEIGHBOUR with no calliper** | 46,840 (1,610) | 54,290 (2,860) | **-7,450 (3,050)** | [-14,140 : 2,090] |
| **GAUSSIAN KERNEL with bandwidth by normal approximation** | 46,840 (1,610) | 54,050 (1,280) | ***-7,210 (1,690)*** | [-11,630 : -4,520] |
| | **OUTCOME: PER CAPITAL NET HOUSEHOLD INCOME (in Deutsch Marks, Base 1999)** | | | |
| | $E[Y^{ADD}\|S=d]$ | $E[Y^{AAA}\|S=d]$ | $\hat{\vartheta}$ (impact estimate) | 95% Confidence Interval on $\hat{\vartheta}$ |
| **NEAREST NEIGHBOUR with no calliper** | 36,660 (800) | 37,600 (1,570) | **-940 (1,685)** | [-6,250 : 380] |
| **GAUSSIAN KERNEL with bandwidth by normal approximation** | 36,660 (800) | 39,060 (690) | ***-2,410 (915)*** | [-4,570 : -930] |

# 7 Conclusions

In this paper we estimate the impact of disability status on three different labour market outcomes (working versus not working, annual labour income and per capita household disposable income), allowing for two different types of underlying populations, one that includes anyone of prime age, SSC1, and a second that restricts the sample to those who, been in prime age, declare to be full time workers at the first period of each three-year sequences in which they might become disabled.

Our empirical section makes use of matching methods to allow for the counterfactual approach associated with treatment effect techniques for program evaluation. In particular, we estimate by matching on the propensity score. Such methods improves on other parametric and semi-parametric approaches to program evaluation because they avoid many potential biases due to model misspecification. At the same time, matching on the propensity score allows to compare the outcome of sub-groups in the same support as defined by a set of observed characteristics.

Our empirical study is based on data from 18 waves of the German Socio-Economic Panel (GSOEP, 1984-2001), thus using all years for which the information is available. This panel provides information not only with respect to all labour market outcomes of interest, but also information on legal disability status, while being also rich on information with respect to other social and economic aspects needed for the analysis. The use of several waves allows us to construct two groups of individuals defined as treated individuals and control (or comparison) individuals. Those in the treatment group are individuals who, been non-disable at a particular year, become disable and remain so in the consecutive second and third year. According to German law, from the moment a person becomes (legally) disable she is entitled to advantages (e.g., particular re-training, free rehabilitation, subsidised wages for employers, etc.) which should help her to lower the cost of engaging in paid labour market activities. Thus, we assume these policies to have some impact on the observed labour market outcome of treatment units, given that they have been disable for at most two years. The control group are individuals who declare non-disability status over a given set of three years, and, therefore, do not receive the impact of policies which are built specifically for disable persons.

We estimate the propensity score using variables grouped according to categories of observed pre-disability characteristics which may have an effect on both labour market outcomes and the probability of becoming disable. Matching on the propensity score can lead to different estimates according to which matching methods is used. We compares estimates using two different matching techniques, namely, Nearest Neighbour and Kernel matching. As expected, the magnitudes of the estimates do not differ much between methods. Where they differ is with respect to sampling variance estimates, with Nearest Neighbour consistently showing larger variance: this is expected since, for any given calliper, the Nearest Neighbour method is based on placing all the weight in one observation in the control group only to minimise bias, whereas the counterfactuals assigned when using Kernel estimation give weight to more than one observation, thus reducing variance in the typical bias-variance trade-off.

With respect to the outcome “working” versus “not working”, results suggest that non-disable fair better in that they might experience as much as a 9.6% greater chance to be working than their disable counterparts. Similarly, after disability has kicked in, there is a considerable earnings gap between the non-disable and the population of disable individuals, with those in disability status experiencing annual earnings differentials of up to approximately DM 6,000, translating into a 16% earnings differential between the two groups. With respect to per capita disposable income, the differential is not so large as with earnings, showing that non-disable enjoy, on average, levels of annual disposable income that are between 1% and 5% higher than the prime age disable population. However, such differential is statistically significant.

We have compared estimation for two sample definitions. Whereas the one commented so far focuses on a selection of the population that affects prime age individuals, we have repeated the exercise focusing on individual who, while been of prime age, declare to be full time workers at the beginning of the period under study (when all individuals are, by construction, non-disable). We believe that, once the disability kicks in, and given that this sub-population is already actively engaged in full time work, the policies will be more relevant and the impact of these policies more prominent so that we might expect the impact of disability to be lower in magnitude (for example, they might become more aware of the different re-training programs, be better advised, etc.). However, estimates of the effect of disability on the labour market outcomes for these individuals who declare to be full time workers at a pre-disability stage do not differ in magnitude (or significance) relative to the impact that disability might have in the overall population of prime age, thus further reinforcing those conclusions obtained under SSC1 with respect to the effectiveness of the aim of disability policies. If at all, and comparing the ATET estimate of this initial full time workers to estimates using the original sample (all prime age individuals), the suggestion is that been in full time employment, once there is a shift from a non-disable to a disable status, the probability of been out of work can increase to range between 8.5% and 9.2%, on average their earnings will be reduced, possibly by as much as 20% relative to their non-disable counterparts, and will experience greater purchasing power inequality than the non-disable population, with their per capita disposable income been reduced possibly as much as 6.1% when compared to those who did not receive a disability shock.

Besides our contribution to a growing area in applied econometrics that uses treatment effect techniques for program evaluation, we believe that this paper makes an important contribution to the understanding of how disable individuals fair in the labour market. Many studies suggest an adverse participation and wage differential for disable individuals who are able and willing to participate in paid labour market activities. For example, in the case of Germany, Albrecht and Braum (1998), suggest large differentials in unemployment rate, whereas the Kid et al. (1998) study estimates a 50% gap in the participation rate, where only half of it can be explained by productivity differentials. Because of our particular way of constructing treatment and control units, a direct implication of our results is that policies aiming at helping disable individuals into paid labour market activities are not effective at their aim, either because they do not motivate individuals sufficiently to take up paid activities (as would show the results of SSC1), or because they do not break down the barriers of entry for those who have become disable but who where previously engaged in full time paid labour market activities.

Furthermore, earning and purchasing power differentials suggest that social differentials between the disable and non-disable exist, in itself something that might act as a deterrent for those who, being disable, might otherwise choose to participate as active members in the labour market. In Germany the system obliges employers with a workforce of 16 or more to employ disable employees at least up to a minimum of 6% of the total workforce. Non fulfilment of this quota is often used to suggest that disable fair worst in the labour market, with less chances of been employed than non-disable. To some extend our work provides quantitative backing for such claims.

## References

Aakvik, A., (2003), "Estimating the employment effects of education for disabled workers in Norway", *Empirical Economics*, vol.28, 3, 515-534.

Albrecht, M., and H. Braun (1998), “International research project on job retention and return to work strategies for disable workers”. Study Report: Germany, ed. *International Labour Organization*, Geneva.

Black, D., and J. Smith (2002), “How Robust is the evidence on the effects of college quality? Evidence from matching”, *University of Maryland working paper*.

Contoyannis, P. and N. Rice (2001), “The impact of health on wages: Evidence from the British Household Panel Study”, *Empirical Economics*, vol.26,4,599-622.

Even, W. E., and D. A. MacPherson (1990), “Plant size in the decline of unionism”, *Economic Letters*, 32, 392-398.

Härdle, W., and J. S. Marron (1985), “Optimal bandwidth selection in nonparametric regression functions estimation”, *Annals of Statistics*, 13, 1465-1481.

Heckman, J. J., and J. Hotz (1989), “Choosing among alternative methods for estimating the impact of social programs: The case of manpower training”, *Journal of the American Statistical Association*, 84, 862-874.

Heckman, J. J., H. Ichimura, and P. Todd (1997), “Matching as an econometric evaluation estimator: evidence from evaluating a job training programme”, *Review of Economics Studies*, 64, 605-654.

Kidd, M., P. Sloane and I. Ferko (2000), “Disability and labour market: an analysis of British males”, *Journal of Health Economics*, 19, 961-981.

Kreider, B., and J. Pepper (2002), “Disability and employment: re-evaluating the evidence in light of reporting errors”, *CRR WP 2002-6*, Boston College.

Lechner, M. (1999), “Continuous off-the job training in East Germany after Unification”, Journal of Business and Economics Statistics, 17, 74-90.

Manski, C., (1994), “The selection problem”, in: *Advances in Econometrics,* ed. C. Sims, Cambridge University Press, 143-170.

Mueser, P., K. Troske and A. Gorislavsky (2003), “Using state administrative data to measure program performance”, *Working paper*, University of Missouri-Columbia.

Oyen, R. (1989), “Vocational training, labour market opportunities, and internal company integration of disable persons. Survey of the empirical research of the 1980s”, *Mitteilungen aus der Arbeitsmrket – und Berufsforschung (MittAB)*, ed. Stuttgart et al., 4, 507-522.

Rosenbaum, P., and D. Rubin (1983), “The central role of the propensity score in observational studies for causal effects”, *Biometrika*, 70(1), 41-55.

Rubin, D. (1974), “Estimating casual effects of treatments in randomised and non-randomised studies”, *Journal of Educational Psychology*, 66, 688-701.

Silverman, R. (1986), “Density estimation”, *London: Chapman and Hall*.

Smith, J. and P. Todd (2002), “Does matching overcome LaLonde’s critique of non-experimental methods?”, *Journal of Econometrics*, forthcoming.

Sundberg, G. (1996), “Essays on Health Economics”, *PhD. Thesis Series*, Department of Economics, Uppsala University.

Thornton, P. (1998), “International research project on job retention and return to work strategies for disabled workers. Key issues”. *International Labour Organization Publications*, Geneva.

Thornton, P. and N. Lunt (1997), “Employment policies for disabled people in eighteen countries: a review”, *York Social Policy Research Unit WP series*, University of York.

Walker, I and A. Thompson, (1996), “Disability, wages and labour force participation: Evidence from UK panel data”, *WP Series Department of Economics, Keele University*, Working Paper no.96/14.

Williamson, J. and T. McNamara, “The effects of unplanned changes in marital and disability status: interrupted trajectories and labour force participation”, *CRR WP 2002-6*, Boston College.

## Appendix A:

Each wave contains a set of questions with reference to health status. Among this questions, individuals are asked if (a) they are classified as legally disable, and (b) the degree (in percentages) of disability, if any. Both questions are reproduced below, in their English translation format.

*(a)* *“Are you officially registered as having a reduced capacity for work or of being severely disabled? (If you are receiving disability benefits, then enter “yes”).*

*Yes________*

*No_________*

*(b)* *“If yes to (a), what is the degree of your disability?*

*percentage of disability________*

The above questions refer to questionnaires from 1984 to 1997. In 1998 and thereafter the wording of both questions changed slightly, although the intended information remained equivalent. The new wording is reproduced below:

*(c)* *“Are you legally classified as handicapped or capable of gainful employment only to a reduced extent due to medical reasons? (If you receive social security due to disability, please enter “yes”)*

*Yes________*

*No_________*

*(d)* *“What is the extend of this capability reduction or handicap according to the most recent diagnosis?*

*_______%*

## Appendix B:

We are interested on estimating the impact of disability on labour market outcomes at $t_3$ of individuals who been non-disable at $t_1$, become disable at $t_2$ and remain classified as disable at $t_3$. Our data is based on an annual survey, meaning that we require individuals who are observed for at least for 3 consecutive years. For example, if the first wave is for the year 1994, we need to observe both the labour market outcome $Y_{t3}$, and the health pattern – non-disability or disability – of the sample consecutively for the years 1994, 1995 and 1996. However, from 1994 onwards the survey provides a total of 7 years of information (1994 to 2000), such that 5 different sequence of three years each can be used to estimate the parameter of interest. The following shows the algorithm to estimate the final ATET:

Step 1: Starting with the first three year sequence $S_j$; $j=1$, select individuals from the original $N$ sample who are, first, observed consistently over the three year period, and secondly, have a health pattern either defined as AAA (or $a$ for short) – the controls or comparison group – or ADD (or $d$ for short) – the treated group. Disregard any other units in the sample. The sample $n$ of controls and unit form a mutually exclusive binary outcome, with $n = n_a + n_d$. Our assumption is that the original $N$ sample is a representative sample of the target population.

Step 2: Select any variable in the information set at $t_1$ within the survey that might be thought to have an effect on both the treatment and outcome of interest. Let the $(k,n)$ matrix $X_1 = [X'_{a,1} : X'_{d,1}]$ identify these variables allowing for any properly justified interaction between them. With this, estimate the propensity scores $p_j(x_{a,1}) = \Phi_j(x'_{a,1}\hat{\beta}_n)$ and $p_j(x_{d,1}) = \Phi_j(x'_{d,1}\hat{\beta}_n)$ for comparison and treatment groups, respectively, where $\mathbf{\Phi}$ stands for the cumulative normal distribution, and $\hat{\beta}_n$ is the parameter estimate of a binary model (e.g., Probit) such that $P(ADD=1 \mid X_1 = x_1; \beta)$.

Step 3: With an appropriate matching method, compare the distance of the $ith$ element of the estimated propensity score vector for treated units $p_j(x_{d,1})$, to all elements in $p_j(x_{a,1})$, the estimated propensity score vector for the comparison units. The $ith$ element in the treated group receives the counterfactual outcome $Y_i^c$, where $Y_i^c$ is the labour market outcome belonging to the comparison unit that minimises $\left|p_j^i(x_{d,1}) - p_j(x_{a,1})\right|$. Repeat the process for each $i$ unit in $n_d$ to end up with a vector of counterfactual outcomes $y_j^c = (y_1^c, \ldots, y_{nd}^c)'$.

Step 4: Repeat step 1 to step 3 for each of the available 3 year sequences. In our case we end up with 5 vectors of counterfactuals $y_1^c, y_2^c, \ldots, y_5^c$, one for each of our constructed 3 year sequence.

Step 5: Estimate the expected value of the counterfactual outcome with the sample average such that $\hat{E}[Y^a \mid ADD] = (1/n_d)\sum_{j=1}^{5}\sum_{l} y^c_{l\in j}$, where $l$ is the number of treatment units in sequence $j$. Do the same with respect to the expected value of the actual outcome for the treated units, such that $\hat{E}[Y^d \mid ADD] = (1/n_d)\sum_{j=1}^{5}\sum_{l} y^d_{l\in j}$. The average treatment effect on the treated, or ATET, is given by $\hat{E}[Y^a \mid ADD] - \hat{E}[Y^d \mid ADD]$.

To find the empirical distribution for each estimated ATET, repeat steps 1 to 5 an appropriate number of times (for example, 500), each time re-sampling with replacement from the original $N$ in the survey (defined as anyone who is observed over each of the three year sequences). This process will give a vector of ATET estimates, $(ATET_1, ..., ATET_{500})$. The standard error or quantiles of the sampling error for the actual estimate ATET is obtained by estimating the standard error of $(ATET_1, ..., ATET_{500})$ and its distribution (Appendix G provides a visual interpretation of the empirical distribution of these estimates using the SSC1 sample, while similar estimate with respect to SSC2 are available from the Authors on request).

## Appendix C: Results from SSC2

Section 3.1 provides several tables for the analysis of the data defined as SSC1. In this appendix we provide similar summary statistics for the sample defined under the sample selection criterion SSC2. This corresponds to a sub-selection from SSC1, where those selected comply with the conditions imposed to SSC1 but furthermore, classify themselves as been full time employed in the first period under observation. [38] Initially 10,995 belong to the SSC1 selection. Of these, 7,611 comply with the sample selection criteria defined under SSC2.

[38] In this set we include anyone that declares to be in maternity leave, since we have no information of whether these individuals were full or part-time employed at the time of leave.

*Table C1: Distribution, for each wave, between Non-disable and Disable, and within group distribution according to degree of disability (annual sample according to sample selection criteria SSC2)*

| | | | | NON-DISABLE (degree of disability between 0 and 29%) | | | DISABLE (degree of disability between 30% and 89%) | | |
|---|---|---|---|---|---|---|---|---|---|
| **Year** | **New Entries** | **Attrition units** | **Net sample size** | **As % of Net Sample size** | **With a degree of disability = 0** | **Disability of 1% to 29%** | **As % of Net sample size** | **Disability between 30% and 49%** | **Disability between 50% and 89%.** |
| **1984** | 3,738 | - | **3,738** | **94.2** | 92.5 | 0.75 | **5.8** | 37.7 | 62.3 |
| **1985** | 471 | 647 | **3,562** | **94.7** | 93.2 | 0.68 | **5.3** | 29.1 | 70.9 |
| **1986** | 356 | 477 | **3,441** | **93.6** | 91.9 | 0.81 | **6.4** | 30.9 | 69.1 |
| **1987** | 273 | 316 | **3,398** | **94.7** | 92.2 | 0.78 | **5.3** | 32.0 | 68.0 |
| **1988** | 230 | 397 | **3,231** | **95.0** | 92.4 | 0.76 | **5.0** | 25.5 | 74.5 |
| **1989** | 261 | 300 | **3,192** | **95.0** | 91.9 | 0.81 | **5.0** | 30.2 | 69.8 |
| **1990** | 250 | 329 | **3,113** | **94.0** | 90.8 | 0.92 | **6.0** | 53.2 | 46.8 |
| **1991** | 246 | 165 | **3,194** | **95.3** | 92.1 | 0.79 | **4.7** | 36.4 | 63.6 |
| **1992** | 230 | 271 | **3,153** | **95.4** | 92.5 | 0.75 | **4.6** | 38.9 | 61.1 |
| **1993** | 208 | 265 | **3,096** | **95.6** | 93.3 | 0.67 | **4.4** | 44.5 | 55.5 |
| **1994** | 216 | 332 | **2,980** | **95.2** | 92.3 | 0.77 | **4.8** | 50.0 | 50.0 |
| **1995** | 183 | 166 | **2,997** | **94.9** | 92.6 | 0.74 | **5.1** | 51.6 | 48.4 |
| **1996** | 183 | 275 | **2,905** | **95.0** | 91.7 | 0.83 | **5.0** | 50.7 | 49.4 |
| **1997** | 171 | 234 | **2,842** | **95.2** | 91.8 | 0.82 | **4.8** | 56.3 | 43.7 |
| **1998** | 164 | 266 | **2,740** | **95.2** | 91.5 | 0.85 | **4.8** | 50.0 | 50.0 |
| **1999** | 162 | 328 | **2,574** | **94.8** | 90.8 | 0.92 | **5.2** | 50.0 | 50.0 |
| **2000** | 138 | 234 | **2,478** | **95.2** | 90.6 | 0.94 | **4.8** | 49.3 | 50.8 |
| **2001** | 131 | 224 | **2,385** | **95.3** | 90.8 | 0.92 | **4.7** | 49.3 | 50.5 |

Note 1: This table is analogous to Table 3, Section 3.1

***Table C2: Summary statistics for two definitions of labour income and per capita household income according to disability status (all estimates are in Deutsch Marks, base year 1999, SSC2)***

| | Labour Income | | NON-ZERO Labour income | | Per Capita household disposable income | |
|---|---|---|---|---|---|---|
| **Year** | **Non-disable** | **Disable** | **Non-disable** | **Disable** | **Non-disable** | **Disable** |
| **1984** | | | | | | |
| **Mean (s.d)** | 52,300 (20,500) | 48,900 (20,400) | 48,340 (34,700) | 49,600 (19,700) | 31,100 (16,100) | 31,000 (11,400) |
| **Median** | 49,900 | 45,700 | 43,000 | 45,900 | 28,900 | 29,600 |
| **IQR/Q50** | 0.57 | 0.48 | 0.55 | 0.49 | 0.50 | 0.46 |
| **1987** | | | | | | |
| **Mean (s.d)** | 49,700 (31,700) | 52,300 (20,500) | 50,600 (31,200) | 52,900 (19,900) | 34,600 (16,900) | 34,400 (10,600) |
| **Median** | 45,100 | 49,900 | 45,600 | 50,600 | 32,200 | 33,000 |
| **IQR/Q50** | 0.57 | 0.42 | 0.56 | 0.41 | 0.50 | 0.37 |
| **1990** | | | | | | |
| **Mean (s.d)** | 55,670 (45,000) | 54,840 (20,900) | 56,400 (44,900) | 55,500 (20,100) | 38,100 (20,800) | 37,600 (12,900) |
| **Median** | 48,900 | 52,600 | 49,200 | 52,700 | 35,100 | 35,100 |
| **IQR/Q50** | 0.56 | 0.41 | 0.56 | 0.40 | 0.51 | 0.52 |
| **1993** | | | | | | |
| **Mean (s.d)** | 59,900 (38,500) | 62,000 (28,200) | 61,400 (37,700) | 63,400 (26,900) | 42,300 (19,900) | 43,200 (19,000) |
| **Median** | 53,700 | 57,200 | 54,600 | 57,900 | 39,000 | 38,800 |
| **IQR/Q50** | 0.62 | 0.40 | 0.58 | 0.39 | 0.52 | 0.49 |
| **1996** | | | | | | |
| **Mean (s.d)** | 61,600 (41,100) | 60,400 (29,400) | 63,800 (40,100) | 61,700 (28,300) | 42,100 (22,200) | 42,900 (16,900) |
| **Median** | 56,500 | 56,800 | 57,300 | 57,100 | 38,300 | 38,400 |
| **IQR/Q50** | 0.62 | 0.42 | 0.60 | 0.42 | 0.52 | 0.58 |
| **1999** | | | | | | |
| **Mean (s.d)** | 64,900 (37,500) | 60,700 (32,300) | 65,600 (37,100) | 61,700 (31,700) | 42,500 (19,900) | 44,100 (16,900) |
| **Median** | 58,800 | 58,800 | 59,000 | 59,400 | 39,300 | 42,100 |
| **IQR/Q50** | 0.58 | 0.38 | 0.57 | 0.39 | 0.52 | 0.43 |
| **2001** | | | | | | |
| **Mean (s.d)** | 68,150 (43,700) | 62,200 (30,100) | 69,100 (43,200) | 63,300 (29,100) | 44,300 (22,600) | 44,700 (15,000) |
| **Median** | 60,700 | 58,700 | 61,100 | 58,700 | 40,500 | 41,200 |
| **IQR/Q50** | 0.58 | 0.61 | 0.57 | 0.51 | 0.53 | 0.43 |

Note 1: IQR/Q50 refers to the relative Inter Quartile Range, weighted by the median. This estimate allows for a measure of inequality, while weighting by the median controls for overall shifts in the income distribution. If IQR/Q50=1, this reflects perfect equality, while IQR/Q50=0 reflects perfect inequality, with $IQR/Q50 \in [0,\infty)$. This measure is used with respect to per capita household income, since this constitutes a measure close to purchasing power inequality.

Note 2: Table C2 is analogous to Table 5 in Section 3.1. Notice that what would be analogous to Table 4 – Section 3.1 – for the SSC2 sample becomes redundant, given that by definition all selected work in the first period, therefore distribution between working and not working in each year considered would not be information of interest.

***Table C3: Distribution of sequential observations between not –used individuals, untreated (control) group and treated group (SSC2)***

| Sequences | [t1, t2, t3] | Total number of sample points | Total number who do not contribute to either control or treatment | Total in the untreated sample, i.e., comparisons [AAA] | Total in the treated sample, i.e., treatments [ADD] |
|---|---|---|---|---|---|
| S1 | [1984, 1985, 1986] | 2,954 | 270 | 2,661 | 23 |
| S2 | [1985, 1986, 1987] | Na | na | na | Na |
| S3 | [1986, 1987, 1988] | 2,856 | 247 | 2,583 | 26 |
| S4 | [1987, 1988, 1989] | 2,731 | 249 | 2,463 | 19 |
| S5 | [1988, 1989, 1990] | 2,713 | 335 | 2,360 | 18 |
| S6 | [1989, 1990, 1991] | 2,661 | 233 | 2,390 | 38 |
| S7 | [1990, 1991, 1992] | Na | na | na | na |
| S8 | [1991, 1992, 1993] | 2,726 | 197 | 2,500 | 29 |
| S9 | [1992, 1993, 1994] | Na | na | na | Na |
| S10 | [1993, 1994, 1995] | 2,661 | 230 | 2,402 | 29 |
| S11 | [1994, 1995, 1996] | 2,602 | 239 | 2,345 | 18 |
| S12 | [1995, 1996, 1997] | 2,572 | 229 | 2,322 | 21 |
| S13 | [1996, 1997, 1998] | 2,443 | 219 | 2,206 | 18 |
| S14 | [1997, 1998, 1999] | 2,354 | 185 | 2,150 | 19 |
| S15 | [1998, 1999, 2000] | 2,267 | 186 | 2,070 | 11 |
| S16 | [1999, 2000, 2001] | 2,097 | 183 | 1,901 | 13 |
| | **Totals** | | | **30,353** | **282** |

Note1: Sequences S2 (1985 to 1987), S7 (1990 to 1992) and S9 (1992 to 1994) are not used in the final estimates of Section 5 since they show zero counts of units with patter ADD over their corresponding 3 year periods. This table is analogous to Table 6 in Section 3.1.

***Table C4: Change over time (between $t_1$ and $t_3$, average over each of the two periods), for outcome variables (income, household income and employment), and two selected subjective measures (health and work satisfaction) (SSC2)***

| | | *Comparison Group [A(t1)A(t2)A(t3)] Sample size = 30,353* | *Treatment Group [A(t1)D(t2)D(t3)] Sample size = 282* | *One sided t-test* |
|---|---|---|---|---|
| **Annual Labour income** | *Increased* | 0.635 | 0.504 | 4.38 |
| | *Stayed the same* | 0.014 | 0.007 | 1.40 |
| | *Decreased* | 0.351 | 0.487 | -4.55 |
| **Annual Per Capita household net income** | *Increased* | 0.602 | 0.557 | 1.52 |
| | *Stayed the same* | - | - | |
| | *Decreased* | 0.398 | 0.443 | -1.51 |
| **Employment status (E= Employed) (NE=Not-Employed)** | *From E to NE* | 0.061 | 0.191 | -5.54 |
| | *Stayed E over time* | 0.939 | 0.809 | 5.54 |
| | *Stayed NE over time* | 0.000 | 0.000 | n.a |
| | *From NE to E* | 0.000 | 0.000 | n.a |
| **Average number of hours working per week** | *Increased* | 0.312 | 0.248 | 2.48 |
| | *Stayed the same* | 0.278 | 0.230 | 1.91 |
| | *Decreased* | 0.411 | 0.521 | -3.68 |
| **Satisfaction with work (Subjective)** | *Increased* | 0.308 | 0.270 | 1.43 |
| | *Stayed the same* | 0.284 | 0.213 | 2.90 |
| | *Decreased* | 0.407 | 0.518 | -3.71 |
| **Satisfaction with health (subjective)** | *Increased* | 0.038 | 0.156 | -5.45 |
| | *Stayed the same* | 0.915 | 0.699 | 7.89 |
| | *Decreased* | 0.047 | 0.145 | -4.67 |

Note: Estimates for subjective satisfaction with work are based the restricted sample of those declaring to be working only. This table is analogous to Table 7, Section 3.1. Column 5 shows a one tail test of significance difference between AAA and ADD for each paired of estimated probabilities. In the cell "Employment Status", changes from NE(t1) to E(t3) and NE(t1) to NE(t3) are zero by construction since all at t1 are E.

*Table C5: Distribution of degree of disabilities*

| Degree of disability | Distribution between degree brackets at $t_1$ | Distribution between degree brackets at $t_2$ | Distribution between degree brackets at $t_3$ |
|---|---|---|---|
| **0 degree** | 84.0% | - | - |
| **0< degree # 9** | 0 | - | - |
| **10 # degree # 19** | 0.4% | - | - |
| **20 # degree # 29** | 15.6% | - | - |
| **30 # degree # 49** | - | 57.4% | 48.6% |
| **50 # degree # 69** | - | 32.3% | 37.9% |
| **70 # degree to 89** | - | 10.3% | 13.5% |

Note: Table is based on 282 treated in SSC2. The percentages are the result of constructing the sample according to degree of disability, therefore, 0% of those in the treated group in the groups below 30% of disability occurs by construction. This table is analogous to Table 8, Section 3.1

## Appendix D:

In this appendix we provide a detail description of all variables in Table 9 (Section 3) that enter the estimation process (i.e., level variables in the set $X_{1,j}$, for all j-sequences). Table D1 provides a description of the three outcome variables. Column 1 is the mnemonic used for the variable. Column 2 provides a description of the variable. Column 3 shows minimum and maximum values for each variable and Column 4 identifies the origin within the original raw data. Column 5 provides some information on the construction. Table D2 shows the description of the variables that enter the set of covariates that allow for the CIA assumption. Appendix E shows how these variables are combined (between level and interaction terms) to estimate the parameters of interest that build up results in Section 5.

*Table D1: Description of Outcome variables.*

| Mnemonic | Description | Minimum, Maximum | Origins | Construction Variable (e.g. based on 1984 and 2001 names) |
|---|---|---|---|---|
| **EMP** | Dummy variable that equals 1 if person is actively employed in paid labour market activities (full time, self employed, part time employment and maternity leave) | 0,1 | P | 1984: AP08<br>2001:RP12 |
| **INC** | Variable generated by the GSEOP data, which imputes zero to all those who declare to have had zero earnings based on labour market activities. Corrected for base year 1999. | 0 to max (continuous variable) | PGEN | 1984:I1111084<br>2001:I1111001 |
| **HINC** | Variable generated by the GSOEP data base. It includes all incomes in the household, together with benefits and transfers. Corrected for base year 1999, with a per capita estimate based on the square root of the household size | 0 to max (continuous variable) | PGEN | 1984:I1111384<br>2001:I1111301 |

*Table E2: Description of Outcome variables.*

| Mnemonic | Description | Minimum, Maximum | Origins | Construction Variable (e.g. based on 1984 and 2001 names) |
|---|---|---|---|---|
| **SPORT** | Dummy to indicate if person practices sport regularly (=1), otherwise (=0) | 0,1 | P | 1984: AP0202<br>2001:RP0303 |
| **MOTI** | Ranking Motivation of individuals from very low to very high, allowing for 4 categories, where we use information on attending social gatherings, interest on politics and other social events participation on volunteer basis. | 0 to 3 | P | 1984:AP0203,AP0204,AP0205,AP0206<br>2001:RP0304, RP0306, RP0305, RP0308, RP0309 |
| **CAPITAL** | Dummy to indicate that person lives in Berlin | 0,1 | PEQ | 1984:L1110184<br>2001:L1110101 |
| **CITY** | Dummy to indicate that person lives in a city, other than Berlin (e.g., Hamburg, Hessen or Bremen) | 0,1 | PEQ | 1984:L1110184<br>2001:L1110101 |
| **NOCITY** | Dummy to indicate that person has been classified in a Land, but no city is mention (e.g., Lives in area of Bavaria) | 0,1 | PEQ | 1984:L1110184<br>2001:L1110101 |
| **URBAN1** | If person lives in an area with less than 5000 inhabitants in the surroundings | 0,1 | G | One file for all: GGKBOU |
| **URBAN2** | If person lives in area with around 5000 to 50,000 inhabitants | 0,1 | G | One file for all: GGKBOU |
| **URBAN3** | If person lives in area with around 50,000 to 100,000 | 0,1 | G | One file for all: GGKBOU |
| **URBAN4** | If person lives in area with more than 100,000. | 0,1 | G | One file for all: GGKBOU |
| **BLOW** | Indicator for blue collar worker, unskilled | 0,1 | P | 1984: AP2801<br>2001:RP4001 |
| **BMED** | Indicator for blue collar worker, semi-skilled | 0,1 | P | 1984:AP2801<br>2001:RP4001 |
| **BHIGH** | Indicator for blue collar worker, skilled | 0,1 | P | 1984:AP2801<br>2001:RP4001 |
| **WLOW** | Indicator for white collar worker, unskilled | 0,1 | P | 1984:AP2804<br>2001:RP4004 |
| **WMED** | Indicator for white collar worker, semi-skilled | 0,1 | P | 1984:AP2804<br>2001:RP4004 |
| **WHIGH** | Indicator for white collar worker, skilled | 0,1 | P | 1984:AP2804<br>2001:RP4004 |
| **CLOW** | Indicator for civil servant, unskilled | 0,1 | P | 1984: AP2805<br>2001:RP4005 |
| **CMED** | Indicator for civil servant, semi-skilled | 0,1 | P | 1984:AP2805<br>2001:RP4005 |
| **CHIGH** | Indicator for civil servant, skilled (e.g., functionary with a managerial position) | 0,1 | P | 1984:AP2805<br>2001:RP4005 |
| **SELF** | If declares to be self-employed, but not farmer | 0,1 | P | 1984: AP2802<br>2001:RP4002A,B,C |
| **FARM** | If the variable indicating self-employed, the individuals declares to be a self-employed farmer | 0,1 | P | 1984:AP2802<br>2001:RP4002A,B,C |
| **SIZEF** | Size of the firm where individual works. This is given in categories and we take the categories as given since they simulate higher number per each increase category (e.g., category 1: less than 20 employees, category 2: up to 200, …,category 3: more than 2000 employees). | 1,2,3,4 | PGEN | 1984:BETR84<br>2001:BETR01 |
| **TIMEF** | Time spent in the firms where presently working. | Years, 0 to max. (continuous) | PGEN | 1984: AERWZEIT<br>2001:RERWZEIT |
| **HOURS** | Number of hours, weekly average declared to be spent on paid labour market activities. | Hours, 0 to max (continuous) | PGEN | 1984:AARTZEIT<br>2001:RARTZEIT |
| **SATIW** | Category given by actual question where individu- | Category, 1 | P | 1984:AP0304 |

| | | | | |
|---|---|---|---|---|
| | als declare how satisfied they are with working life, from no satisfaction (1) to very satisfied (10) | to 10, taken as such. | | 2001:RP0201 |
| **AGE** | Age of individual | Continuous variable artificially trimmed to be between 17 and 60. | P | 1984: AP62Z<br>2001:RP13002 |
| **MALE** | Dummy for male, gender | 0,1 | P | 1984:AP57<br>2001:RP13001 |
| **HSIZE** | Household size of individual, including kids | 1 to max, continuous variable | PEQ | 1984:D1110684<br>2001:D1110601 |
| **PARTNER** | Dummy if individual has partner living together under same household. | 0,1 | P | 1984:AP58<br>2001:RP13001 |
| **EDU** | Years of education. | 7 to max, continuous variable | PGEN | 1984:ABILZEIT<br>2001:RBILZEIT |
| **OWN** | Dummy to indicate if person owns house where they live, or owns housing property at all. | 0,1 | H | 1984:AH18,AH39<br>2001:RH06,RH22 |
| **EFLOW** | Dummy to indicate that Father of interviewed individual has achieved a low level of education/ work experience. | 0,1 | BIOPAR EN | VSBIL<br>VBBIL, for all years in one file (Bioparen) |
| **EFMED** | Dummy to indicate that Father of interviewed individual has achieved a medium level of educa- tion/ work experience. | 0,1 | BIOPAR EN | VSBIL<br>VBBIL, for all years in one file (Bioparen) |
| **EFHIGH** | Dummy to indicate that Father of interviewed individual has achieved a high level of education/ work experience. | 0,1 | BIOPAR EN | VSBIL<br>VBBIL, for all years in one file (Bioparen) |
| **EMLOW** | Dummy to indicate that Mother of interviewed individual has achieved a low level of education/ work experience. | 0,1 | BIOPAR EN | MSBIL<br>MBBIL, for all years in one file (Bioparen) |
| **EMMED** | Dummy to indicate that Mother of interviewed individual has achieved a medium level of educa- tion/ work experience. | 0,1 | BIOPAR EN | MSBIL<br>MBBIL, for all years in one file (Bioparen) |
| **EMHIGH** | Dummy to indicate that Mother of interviewed individual has achieved a high level of education/ work experience. | 0,1 | BIOPAR EN | MSBIL<br>MBBIL, for all years in one file (Bioparen) |
| **WEST** | Dummy to indicate that individual's origins are from a western type o society (with similar food costumes, traffic costumes, medical services, etc). These are North America, E.U (West Europe), Australia and New Zealand. | 01, | FF | One file only, PPFAD, and we take CORIGIN, from which we define dummy |

Note1: **P:** Personal level files [ap],….,[rp]. **FF:** PFADD file with general household information for all individuals who ever participated in the survey. **G:** GGKOU file. This file is unique to the 100% release of the GSOEP data set, and contains information on degree of urbanization at household level. **PGEN:** Files that on an yearly basis [apgen],…,[rpgen], gathers generated variables – by the data analysts at the GSOEP centre – and allow for information such as Education and income variables. **PEQ:** Files with variables, also generated from personal and household variables, per year so that the original files are [apequiv],…,[rpequiv]. **H:** Household level files [ah],…,[rh]. **BIOPAREN:** Information, for anyone that ever participated in the panel, on parental background, mostly level of education and/or work experience, origins, mortality, etc.

Note2: Since we are using 18 data sets, Column 5 cannot include all variables names of origin. As guide we include those in 1984 and 2001, and refer to http:/panel.Gsoep.de/soepinfo2001/ for further detail, where a list for all years is available by simply typing the name of any of the two years provided.

## Appendix E: Probit estimates

Section 4 explains the use of a Probit set up to estimate $\hat{\beta}_j$ for each of the sequences of three time periods years, so that $\hat{\beta}_j$ is obtained by estimation of $P_{nj}(ADD=1 \mid X_1;\beta)$. The process requires the choice of variables for the $X_{1,j}$ set. Initially, variables as defined in Table 9 (Section 3) and Appendix D are selected, together with plausible interaction terms. For each sequence $j=1,\ldots,16$, and *each* of this variables, we run the probit set up and use a likelihood ration test as a goodness of fit test. This decides which variables and interaction terms are to enter the final set $X_{1,j}$. With this set, a probit is run so that the parameter of interest is found, and thus the propensity score. Using **Sequence 1 as example (i.e. the sequence defining controls according to their health status over 1984-1986), Table E2** shows the estimation of the overall probit. The selection contains 64 variables plus a constant term.

***Table E2: Probit estimate, P(ADD=1|X), n=5,119 observations.***

| Variable | Coefficient | Standard Error |
|---|---|---|
| Constant | -2.98 | 2.18 |
| Sport | -0.073 | 1.05 |
| Motivation | -0.34 | 0.81 |
| Capital | -1.35 | 0.862 |
| Nocity | 0.25 | 0.262 |
| Urban3 | -0.331 | 0.285 |
| Urban1 | 0.024 | 0.211 |
| Blow | 0.393 | 0.318 |
| Bmed | -0.616 | 0.404 |
| Farm | -1.23 | 1.454 |
| Whigh | -0.05 | 0.471 |
| Chigh | 0.28 | 0.478 |
| Sizef | 0.136 | 0.293 |
| Timef | 0.041 | 0.054 |
| Hours | -0.006 | 0.025 |
| ***Satiw*** | ***-0.322*** | ***0.182*** |
| Age | 0.057 | 0.077 |
| ***Male*** | ***2.878*** | ***0.644*** |
| Hsize | -0.097 | 0.199 |
| Partner | -0.538 | 1.027 |
| Edu | 0.024 | 0.142 |
| Own | 0.119 | 0.154 |
| ***Eflow*** | ***-3.74*** | ***1.84*** |
| ***Emhigh*** | ***-2.17*** | ***0.460*** |
| West | 0.353 | 0.450 |
| Age.Age | 0.0001 | 0.001 |
| ***Age.Sport*** | ***-0.071*** | ***0.024*** |
| Age.Moti | 0.005 | 0.013 |
| ***Age.Sizef*** | ***0.005*** | ***0.003*** |
| Age.Timef | 0.001 | 0.006 |
| Age.Hours | -0.002 | 0.001 |
| Age.Edu | 0.0001 | 0.0001 |
| Age.Partner | 0.0001 | 0.002 |
| Age.Eflow | 0.007 | 0.019 |
| ***Male.Sport*** | ***-0.052*** | ***0.033*** |
| Male.Edu | -0.292 | 0.450 |
| ***Male.Blow*** | ***-0.221*** | ***0.048*** |
| Male.Farm | -0.440 | 0.393 |
| Sport.Moti | 0.63 | 0.386 |
| ***Sport.Satiw*** | ***-3.58*** | ***0.876*** |
| ***Sport.Bhigh*** | ***0.197*** | ***0.076*** |
| Sport.Whigh | -0.481 | 0.501 |
| Sport.Chigh | 0.147 | 0.641 |
| Sport.Hours | -0.112 | 0.652 |
| Sport.Hsize | 0.012 | 0.015 |
| Sport.Edu | -0.231 | 0.171 |
| ***Sport.Satil*** | ***0.094*** | ***0.055*** |
| Moti.Capital | -0.062 | 0.073 |
| ***Moti.Self*** | ***1.051*** | ***0.358*** |
| ***Moti.Hours*** | ***0.72*** | ***0.308*** |
| Moti.hsize | -0.011 | 0.007 |

| | | |
|---|---|---|
| Moti.Partner | 0.045 | 0.111 |
| Moti.Edu | 0.074 | 0.383 |
| Moti.Eflow | 0.016 | 0.044 |
| Moti.Satiw | -0.454 | 0.296 |
| Satiw.Capital | 0.057 | 0.039 |
| Satiw.Urban1 | -0.052 | 0.059 |
| ***Satiw.Self*** | ***0.061*** | ***0.020*** |
| ***Satiw.Sizef*** | ***-0.213*** | ***0.057*** |
| Satiw.timef | -0.013 | 0.022 |
| ***Satiw.hours*** | ***0.005*** | ***0.002*** |
| Satiw.kids | -0.00001 | 0.002 |
| Satiw.edu | 0.018 | 0.017 |
| Satiw.Satil | -0.004 | 0.007 |

Value of the **Likelihood function =** -201.813. **Pseudo-R2 =** 0.243, LR = 129.4, against the restricted model where Chi-Square critical value with 64 degrees of freedom is **CH(df=64)** 47.5, so that the restricted model is rejected. Those variables in ***black*** marking are significant, at least at a 0.10 level.

## Appendix F: Common support (SSC1)

Figure F1: Density, Propensity-scores, S1 [1984-1986]

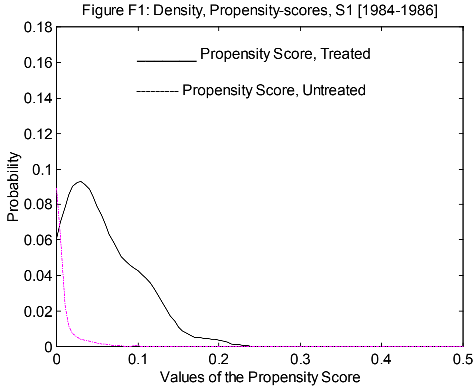



Figure F2: Density, Propensity-scores, S3 [1986-1989]

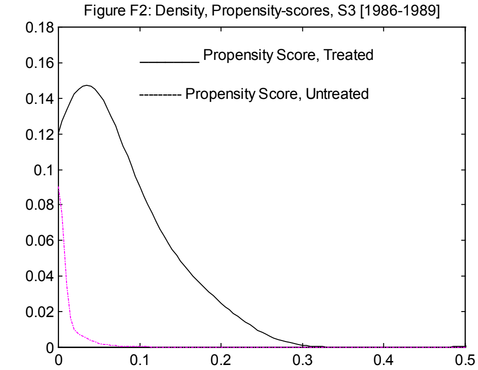



Figure F3: Density, Propensity-scores, S4 [1987-1989]

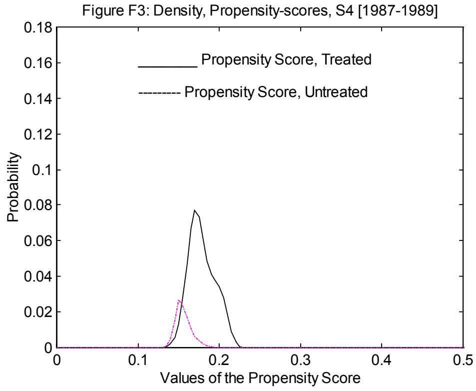



Figure F4: Density, Propensity-scores, S5 [1988-1990]

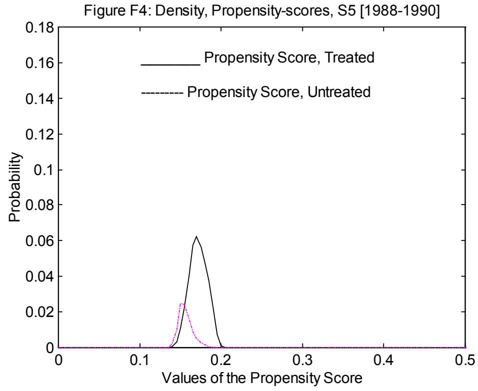


Figure F5: Density, Propensity-scores, S6 [1989-1991]

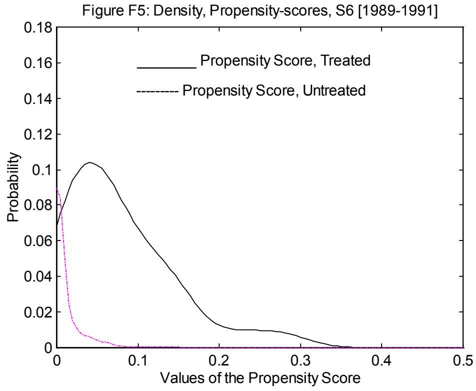



Figure F6: Density, Propensity-scores, S8 [1991-1993]

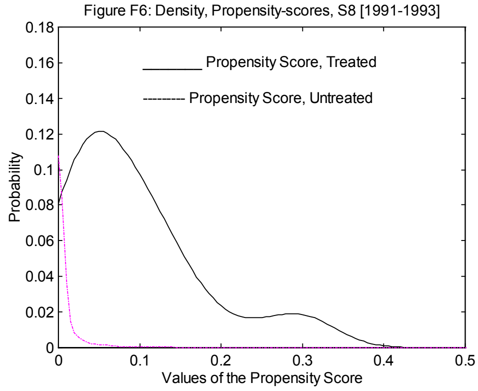



Figure F7: Density, Propensity-scores, S10 [1993-1995]

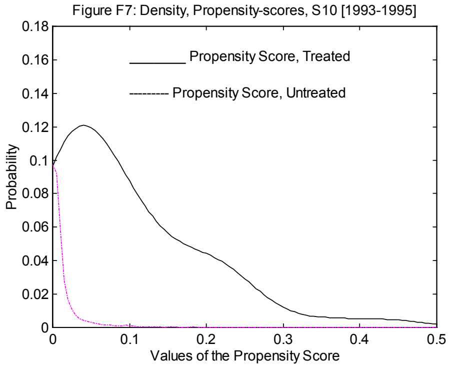



Figure F8: Density, Propensity-scores, S11 [1994-1996]

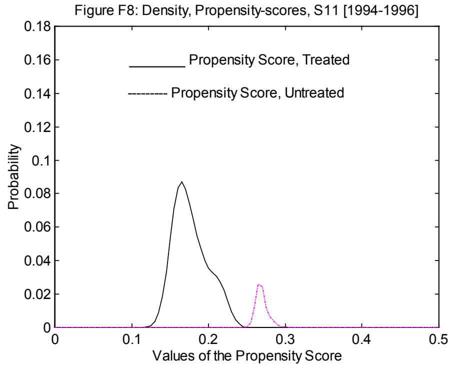


Figure F9: Density, Propensity-scores, S12 [1995-1997]

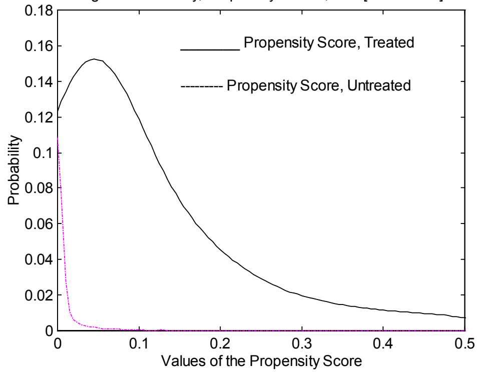



Figure F10: Density, Propensity-scores, S13 [1996-1998]

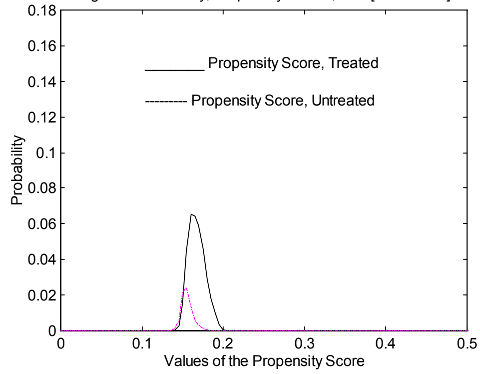



Figure F11: Density, Propensity-scores, S14 [1997-1999]

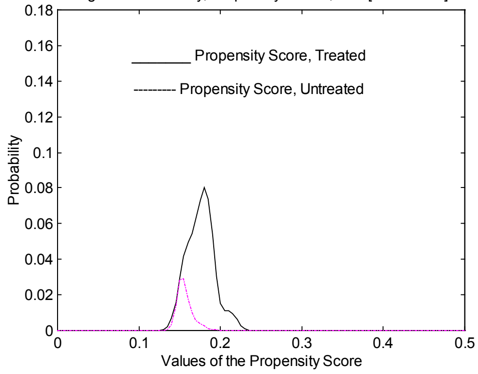



Figure F12: Density, Propensity-scores, S15 [1998-2000]

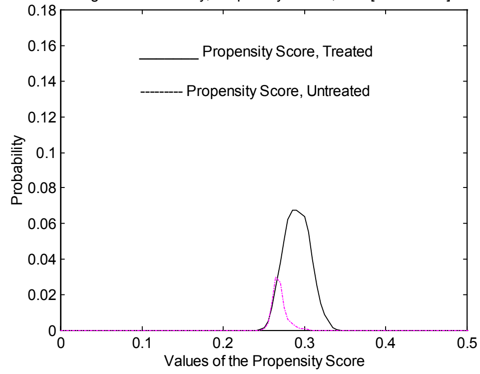
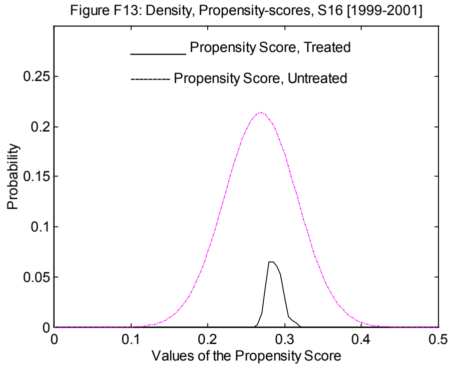



Figure F13: Density, Propensity-scores, S16 [1999-2001]




## Appendix G: Details of the ATET from the bootstrap (SSC1)

Density Estimate of the Average Treatment Effect on the Treated (ATET), based on the empirical distribution of 500 bootstrap estimates. Each Figure shows the density estimate for each of the 7 methods employed to obtain the ATET, and for each of the three outcomes.[39] In each Figure, the density is

[39] For all Figures, NN0 = Nearest Neighbour, no caliper, NN5 = Nearest Neighbour, Caliper = 0.05, NN1 = Nearest Neighbour, Caliper =0.10, KG1 = Kernel Gaussian with a normal approximation bandwidth, KG2 = Kernel Gaussian with bandwidth based on Cross Validation, KE1 = Epanechnikov Kernel with a normal approximation bandwidth and KE2= Epanechnikov Kernel based on Cross validation bandwidth.

compared to the position of the actual ATET estimate (based on the actual sample) by drawing a vertical line over the point estimate.

*Figure G1: Densities for Outcome 'Working'*

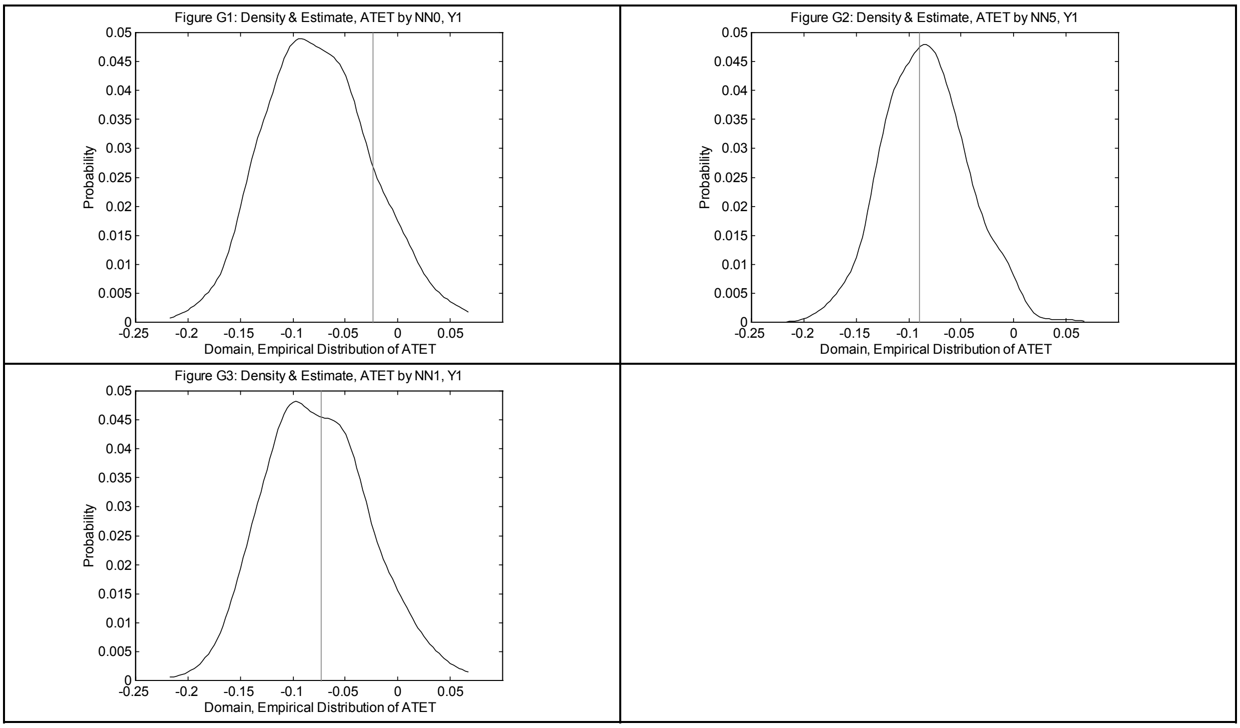


Figure G4: Density & Estimate, ATET by G1, Y1

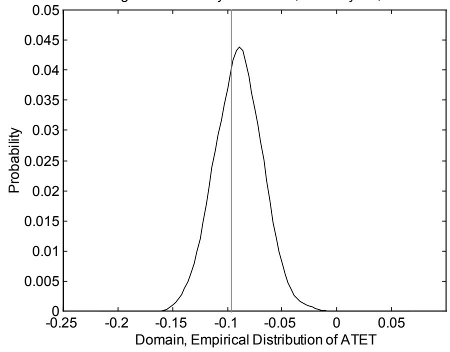



Figure G5: Density & Estimate, ATET by G2, Y1

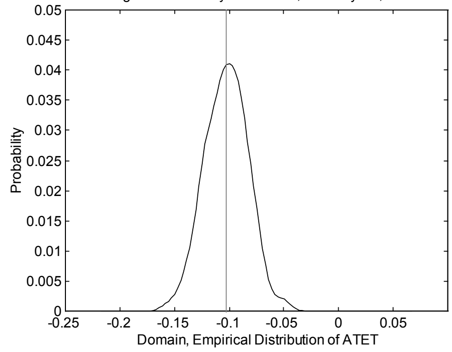



Figure G6: Density & Estimate, ATET by E1, Y1

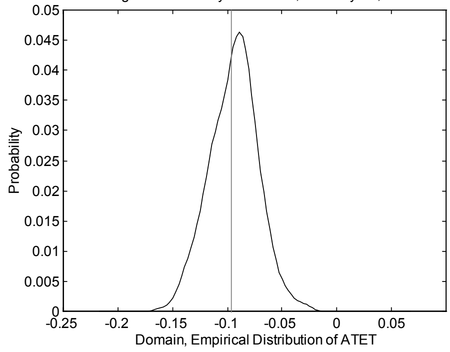



Figure G7: Density & Estimate, ATET by E2, Y1

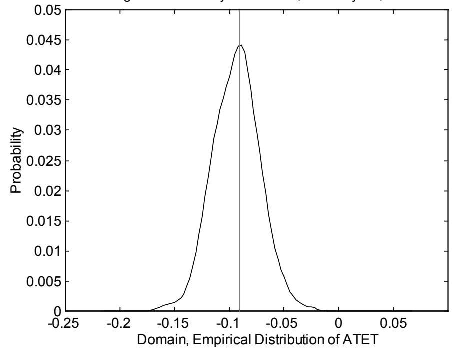


*Figure G2: Densities for Outcome 'Annual Labour Income'*

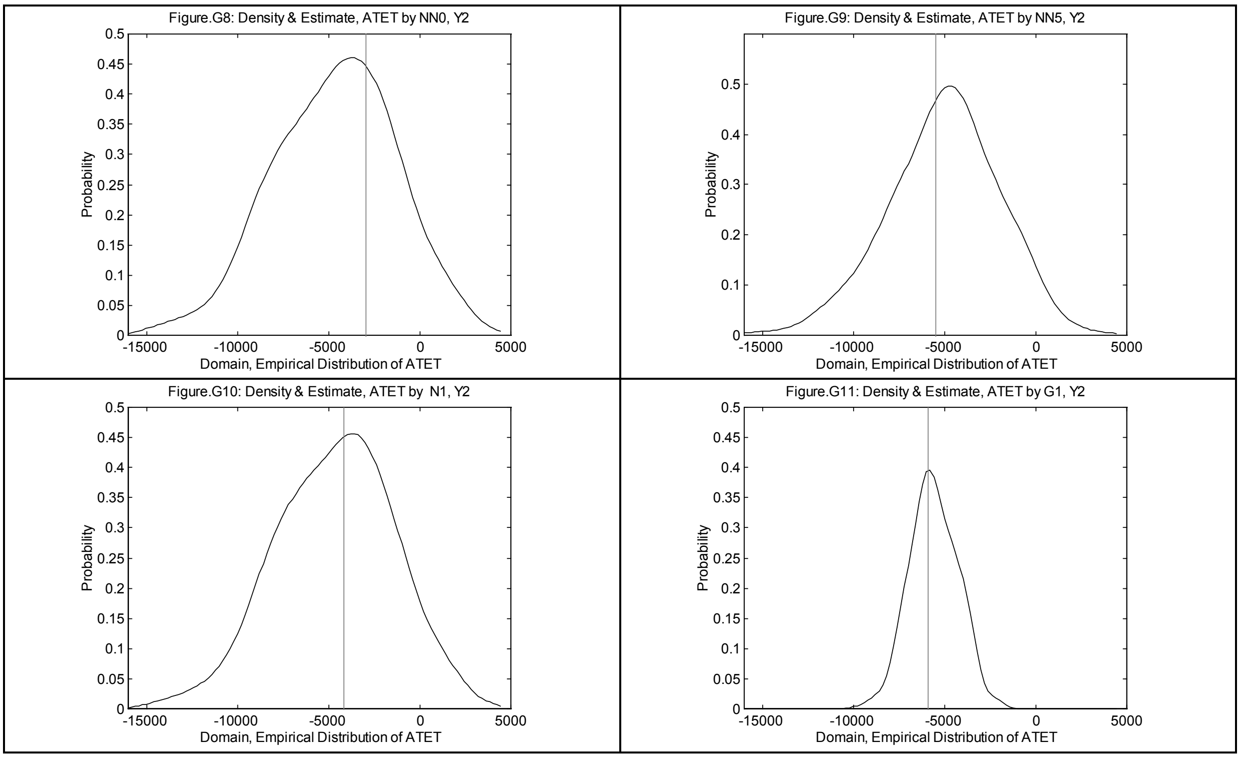


Figure.G12: Density & Estimate, ATET by G2, Y2

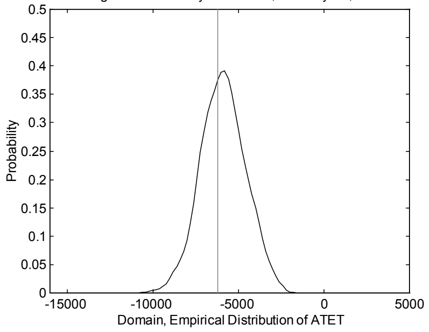



Figure.G13: Density & Estimate, ATET by E1, Y2

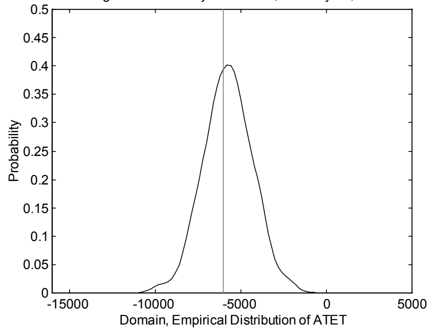



Figure.G14: Density & Estimate, ATET by E2, Y2

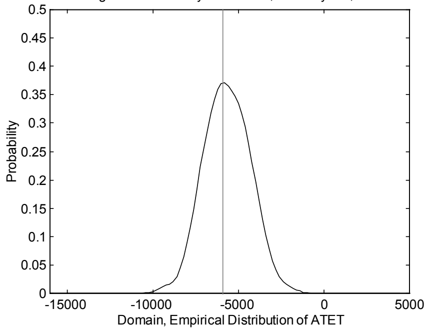


*Figure G.3: Densities for Outcome 'Per Capita household net annual income'*

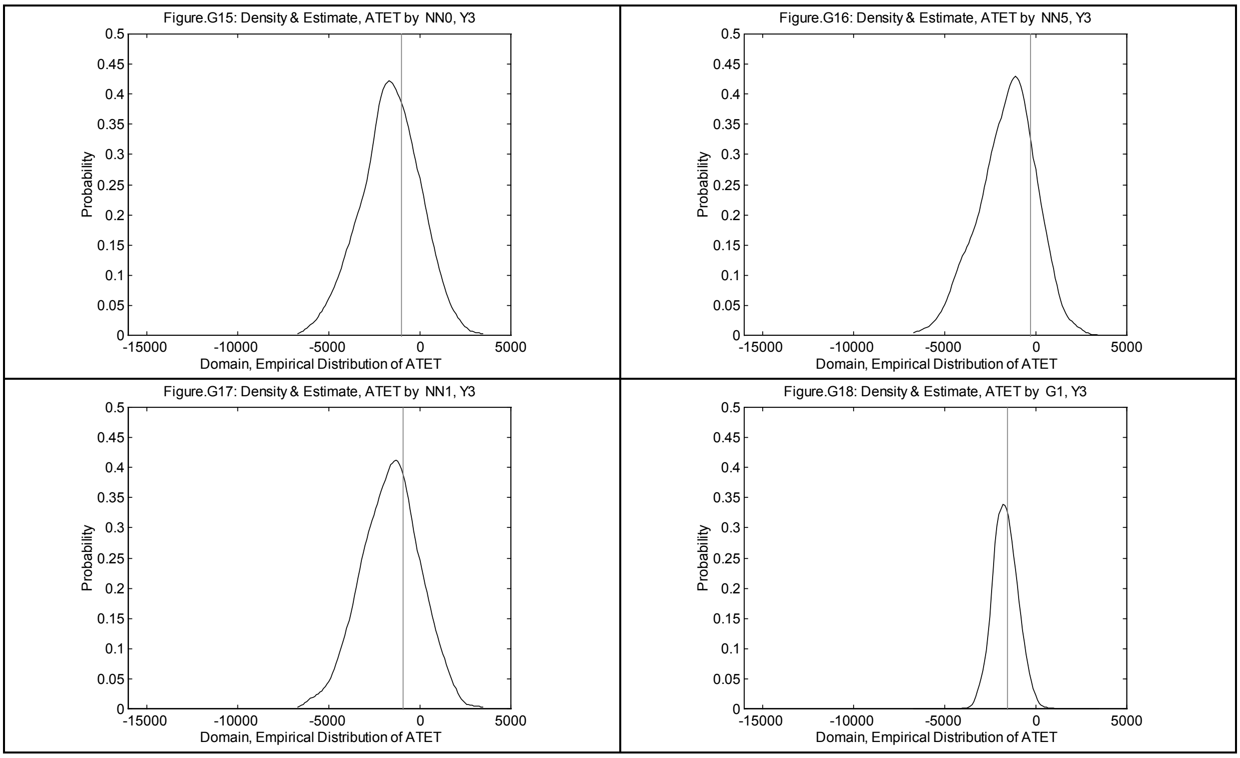


Figure.G19: Density & Estimate, ATET by G2, Y3

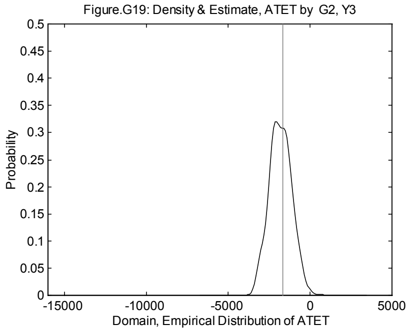



Figure.G20: Density & Estimate, ATET by E1, Y3

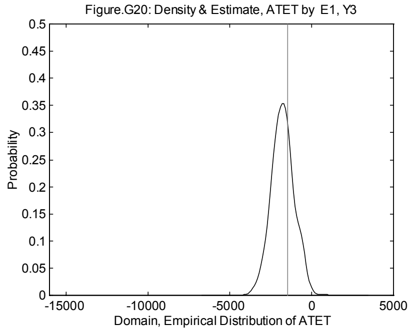
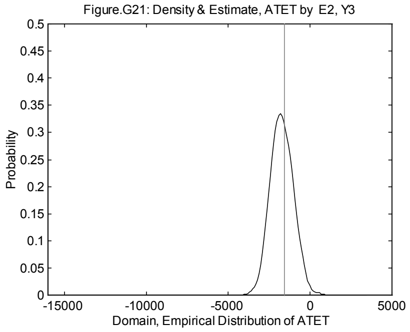



Figure.G21: Density & Estimate, ATET by E2, Y3




## Appendix H: Bandwidth for the Kernel-Method estimation

Gaussian Kernel requires a choice of a bandwidth estimated using the generic form $h_n = h_0 \times \sigma(x) \times n^{(-0.2)}$, where the choice of $h_0$ is some optimal normalizing value, $\sigma(.)$ stands for the standard deviation of the weighting variable, and $n$ is the number of units defined within $x$, the weighting variable. In our case, the bandwidth is with respect to the density of the propensity score for the control group $(p(x_c))$, relative to its distance with each $i^{th}$ unit in the estimated propensity score for the treated group $(p(x_t))$. To estimate we use two different choices for $h_0$. Kernel based estimates of the ATET parameter in Table 12 (results section) are based on a bandwidth by normal approximation following Silverman, such that $h_n = 1.06 \times \sigma(p(x_c)) \times {n_c}^{(-0.2)}$. Table H1 shows the pairs ($h_0, h_n$) for each of the 13 sequences that enter the estimation method.

*Table H1: Bandwidths estimates by normal approximation and Cross-Validation.*

| | SSC1 | SSC2 |
|---|---|---|
| | **Bandwidth by normal approximation (h,h0)** | **Bandwidth by normal approximation (h,h0)** |
| **S1** | 0.004 (1.06) | 0.003 (1.06) |
| **S3** | 0.005 (1.06) | 0.002 (1.06) |
| **S4** | 0.002 (1.06) | 0.002 (1.06) |
| **S5** | 0.002 (1.06) | 0.002 (1.06) |
| **S6** | 0.006 (1.06) | 0.004 (1.06) |
| **S8** | 0.005 (1.06) | 0.003 (1.06) |
| **S10** | 0.006 (1.06) | 0.003 (1.06) |
| **S11** | 0.002 (1.06) | 0.003 (1.06) |
| **S12** | 0.005 (1.06) | 0.007 (1.06) |
| **S13** | 0.001 (1.06) | 0.002 (1.06) |
| **S14** | 0.002 (1.06) | 0.003 (1.06) |
| **S15** | 0.002 (1.06) | 0.002 (1.06) |
| **S16** | 0.050 (1.06) | 0.008 (1.06) |