

Lietzmann, Torsten; Schmelzer, Paul; Wiemers, Jürgen

Article — Published Version

Marginal employment for welfare recipients: stepping stone or obstacle?

Labour

Provided in Cooperation with:

German Institute for Economic Research (DIW Berlin)

Suggested Citation: Lietzmann, Torsten; Schmelzer, Paul; Wiemers, Jürgen (2017) : Marginal employment for welfare recipients: stepping stone or obstacle?, Labour, ISSN 1121-7801, Wiley, Hoboken, Vol. 31, Iss. 4, pp. 394-414,
<https://doi.org/10.1111/lab.12098> ,
<https://onlinelibrary.wiley.com/doi/full/10.1111/lab.12098>

This Version is available at:

<https://hdl.handle.net/10419/200398>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by-nc-nd/4.0/>

Marginal employment for welfare recipients: stepping stone or obstacle?

Torsten Lietzmann¹ — Paul Schmelzer² — Jürgen Wiemers¹

Abstract. Marginal employment (ME) is one of the largest forms of atypical employment in Germany. We analyse whether ME has a ‘stepping stone’ function for unemployed individuals, i.e., whether ME increases the subsequent probability of regular employment. We find differing treatment effects by unemployment duration. According to our results, ME increases the likelihood of regular employment within a 3-year observation period only for those who take up ME several months after beginning to receive benefits. In contrast, for those starting ME within the first months of receiving benefits, there is no effect on the probability of regular employment. Although we took several measures to minimize the impact of unobserved heterogeneity, our results can only be interpreted as causal if the conditional independence assumption holds.

1. Introduction

Throughout Europe, atypical employment is rising (Hipp *et al.*, 2015). In Germany, marginal employment (ME) has been increasing since it was substantively reformed in 2003 and is now one of the most common forms of atypical employment. In 2014, about 7.5 million men and women were marginally employed, i.e., employed with low earnings up to a threshold of 450 € per month that are not subject to social security contributions.¹ As this segment of the labour market grows, we ask questions about its goals and functions.

This study analyses whether ME has a ‘stepping stone’ function for unemployed individuals, i.e., whether ME increases the subsequent probability of regular employment. We add our results to existing empirical studies that have yielded no conclusive evidence for ME’s ‘stepping stone’ function for Germany’s unemployed. Studies that analyse the ‘stepping stone’ function either apply matching techniques to compare the labour market outcomes of initially unemployed individuals who take up ME to the hypothetical counterfactual situation in which these persons remain unemployed or they apply the timing-of-events methodology.

Freier and Steiner (2007, 2008) find that for the period of 1993–2003, taking up a marginal job had no effect on the probability of subsequent regular employment. Nonetheless, they find that marginally employed men subsequently spent less time in unemployment and have higher cumulative incomes. Caliendo *et al.* (2012, 2016) also do not find a positive effect of ME on regular employment for short-term unemployed men in the period

¹Institute for Employment Research (IAB), Regensburger Str. 104, 90478 Nürnberg, Germany.
E-mail: torsten.lietzmann@iab.de

²German Institute for Economic Research (DIW), Mohrenstr. 58, 10117 Berlin.

The copyright line for this article was changed on 25 May 2017 after original online publication.

2001 to 2004. For long-term unemployed men, however, the probability of regular employment is higher for those who did take up a marginal job. In addition, Caliendo *et al.* (2012, 2016) find that those with a preceding temporal marginal job are more likely to find stable regular employment. Lehmer (2012), who focuses on the effects of temporary agency work on long-term unemployed men and women, also provides results for the effects of ME for the period 2004 to 2008, finding that ME has a small positive effect for men but no effect for women.

Our study adds to the literature in the following ways. First, it looks at a more recent time period, i.e., after Germany's major labour market reforms (Hartz reforms) at the beginning of the 2000s. These reforms further improved the attractiveness of ME and introduced a new means-tested social benefit system called 'unemployment benefit II (UB II)' that focuses on labour market activation.

Second, we use a new administrative data source for the recipients of UB II. Unlike the data used in previous studies, our data set includes information on desired labour supply, which we use to restrict the sample to those individuals who are searching for full-time employment. Also, in contrast to previous studies, our data include reliable information on household composition, allowing us to further restrict the sample to childless single men and women, the latter of which have often been excluded because of the impossibility of controlling for the household context. These sample restrictions reduce unobserved heterogeneity as individuals with children and/or in a relationship might be less prone for active job search for regular employment than single and childless men and women, e.g. because of family obligations.

Third, we follow methodological developments in the evaluation literature by applying a dynamic evaluation approach (e.g., Gebel, 2013; Hujer and Thomsen, 2010; Kvasnicka, 2009; Sianesi, 2004; Voßemer and Schuck, 2016) that has not previously been used to analyse ME. The results show the importance of the dynamic approach: We find differing treatment effects by unemployment duration. According to our results, ME does increase the likelihood of regular employment within a 3-year observation period only for those who take up ME several months after beginning to receive benefits. For those starting ME within the first months of receiving benefits, there is no effect on the probability of regular employment.

Note that our results can only be interpreted as causal if we invoke the conditional independence assumption (CIA) because we cannot exploit a quasi-experimental situation and our methodological approach cannot control for possible unobserved heterogeneity in the selection process. However, we took several measures to minimize the impact of unobserved heterogeneity: We (1) apply a dynamic matching approach that entails a weaker CIA assumption than static matching and we employ controls for unemployment duration that should capture further unobserved differences among individuals. In addition, we (2) use data that contain rich information about individual employment histories and (3) select a homogeneous group of childless single men and women searching for full-time employment. Finally, we (4) check the robustness of our results with respect to unobserved heterogeneity by applying the Mantel–Haenszel test.

2. Institutional background

Because our data refer to the period from 2005 to 2009, we describe the institutional regulations for ME in Germany applicable at that time. Thus, we focus on the regulations

introduced with the most recent major ME reform in 2003 (Caliendo and Wrohlich, 2010). This reform aimed to improve labour market flexibility and create more incentives for low-wage jobs (Jacobi and Kluve, 2007). ME is designed to be attractive to low-wage employees because the employees pay no taxes or social security contributions up to a maximum labour income, which until 2012 was 400 € per month.² Since 2003, ME has not restricted the number of weekly working hours.³ It is also possible to hold two or more jobs and still be considered marginally employed provided the combined income from these jobs does not exceed the monthly maximum labour income. Finally, regular employment — i.e., employment subject to taxes and social security contributions — can be combined with one secondary ME job.⁴ In that case, taxes and social security contributions must only be paid for the primary job. In contrast, employers must pay a relatively larger share of social security contributions for ME (31 per cent of gross labour earnings) than for regular employment subject to social security (19 per cent). Otherwise, ME is subject to the same regulations as regular employment, e.g., dismissal protection, entitlements to holidays, continued remuneration in case of sickness, and pay for public holidays.

The design of the ME regulations leads to a special attractiveness of ME for groups that can draw on other sources of private income within the household or the family or who have access to social security entitlements based on some other status; those groups include not only the unemployed but also pensioners, students, housewives, and -husbands (Bäcker and Neuffer, 2012; Körner *et al.*, 2013). Unemployed individuals with marginal jobs also have access to health care via the unemployment benefit system, and their earnings are supplemented by benefits if they pass a means test.

In 2005, Germany's social security system for the unemployed experienced major changes with the introduction of the so-called Hartz IV reform (Eichhorst *et al.*, 2010), which arguably further increased the attractiveness of ME to the unemployed. In this reform, social and unemployment assistance were combined to form the new UB II, which provides basic income support for job seekers and their families if their total household income is insufficient to meet a minimal standard of living and they have no or no sufficient entitlement to unemployment insurance benefits. Along with this new benefit system, a stronger focus on activating the unemployed was included in German labour market policy so that the unemployed should be more willing to accept low-quality jobs. A refusal to accept job offers or a lacking willingness for job search activities can lead to benefit cuts. Compared to the previous systems of social and unemployment assistance, the amount of earned income that benefit recipients are allowed to retain is higher under UB II (Bruckmeier and Wiemers, 2011). Recipients can earn a gross income of 100 € per month before their welfare benefits are reduced. For earnings above 100 € per month, the benefit reduction rate is 80 per cent. Above 800 € per month, that rate is increased to 90 per cent. Earnings above a threshold of 1,200 € (1,500 € for recipients with children) per month reduce the benefits at a rate of 100 per cent. For example, a marginally employed benefit recipient with a monthly earned income of 400 € can retain 160 € of his labour income. The high marginal benefit reduction rates between 80 and 100 per cent create low monetary incentives for marginally employed UB-II recipients to extend their working hours and earnings and to leave ME in favour of regular employment.

For marginally employed workers, a maximum labour income of 400 € is insufficient to secure a minimum standard of living as defined by either UB II or commonly applied poverty thresholds. Furthermore, the exemption from social security contributions is accompanied by a lack of entitlements to health and unemployment insurance. In addition, people in ME have less access to benefits (such as paid vacation, continued remuneration in case

of sickness, and pay for public holidays) than do workers in employment subject to social security contributions (RWI, 2012; Stegmaier *et al.*, 2015). Finally, ME also yields negative long-term risks, missing pension entitlements, and an associated higher probability of poverty in old age. Therefore, for UB-II recipients, it is of particular interest to know whether ME serves as a stepping stone to regular employment or whether it results in a long-term 'lock-in' of employees in a situation of low labour income topped up with UB-II benefits.

3. Theoretical considerations

The effect of taking up ME on the probability of subsequent regular employment is theoretically undetermined.⁵ On the one hand, ME can improve the employment prospects of formerly unemployed men and women because it may decrease human capital depreciation in times of unemployment (Albrecht *et al.*, 1999; Becker, 1975; Gregory and Jukes, 2001; Mincer, 1974). ME might also provide the opportunity not only to increase a worker's general and/or specific human capital through on-the-job training but also to alleviate negative consequences of unemployment, for example, on (mental) health (Jahoda, 1982; Paul and Batinic, 2010). Furthermore, employers might use ME as a screening device for productivity. Finally, the marginally employed might be able to extend their social network and establish contacts that provide them with information about job openings and assist them in their job search (Granovetter, 1973) either inside or outside of the firm. These aspects should increase marginally employed individuals' likelihood of finding regular employment compared to unemployed individuals.

On the other hand, there are reasons that individuals receiving unemployment or social benefits and working in a marginal job are less likely to take up regular employment. First, the marginally employed should have higher reservation wages than the unemployed without ME, and the time that they can devote to a job search should be lower because of the time spent in ME (i.e., the 'lock-in effect' of taking up ME). Second, ME is concentrated in a few sectors of the economy, which might limit both the scope of employment opportunities and the transferability of gained human capital. Third, ME largely consists of low-qualified tasks, which should limit human capital gains.

For these reasons, the question whether ME provides a 'stepping stone' function for the unemployed must be decided empirically. In addition, the existence of an ME 'stepping stone' function is an empirically important question because the number of marginally employed UB-II recipients is quite substantial, with annual averages between 450,000 and 500,000 in recent years⁶ (Statistik der Bundesagentur für Arbeit, 2011).

Moreover, the relative importance of the mechanisms for and against a 'stepping stone' function will likely change with the time of treatment. If ME is taken up relatively early after the beginning of an unemployment spell, labour market prospects should still be comparatively good. Therefore, arguments against a 'stepping stone' effect (higher reservation wages, lock-in effect) should have a relatively high (negative) impact. With an increasing duration of unemployment, however, labour market prospects should decrease. Therefore, the arguments for a 'stepping stone' effect of taking up ME (slowing down the processes of human capital deterioration and declining health) become more relevant if ME is taken up relatively late after entering unemployment. In contrast, the impact of the mechanisms which work against the 'stepping stone' effect of ME should be relatively constant over the duration of unemployment. Summing up, the later ME is taken up in the

unemployment spell, the higher the net effect of taking up ME on subsequent regular employment should be.

4. Data and methodology

Evaluating the effect of ME on the subsequent probability of taking up regular employment by individuals who were formerly unemployed requires longitudinal data that cover labour market histories (unemployment, benefit receipt, employment) for an extended time. To select a sample of unemployed people who are actually at risk of transitioning from ME to regular employment, it is necessary to use a data set that includes information about the extent of the individual labour supply, the existence of a desire to be regularly employed and possible competing obligations and roles to which unemployed people can allocate their time instead of employment and job search (i.e., family obligations). We use administrative data (Administrative Panel SGB II) provided by the German Federal Employment Agency (Rudolph *et al.*, 2013), which is a 10 per cent sample of UB-II recipients in Germany. Our data include longitudinal information about UB-II receipt and about unemployed people's household context. The latter information is important for the means test and the benefit level and was previously unavailable because it has only been gathered reliably since the introduction of UB II. These data are enriched by longitudinal information on whether the unemployed are actually looking for full-time employment from a different administrative data source (Job-Search Histories data; Köhler, 2015) and on times spent in marginal and regular employment along with times spent in measures of active labour market policy (from Integrated Employment Biographies; Jacobebbinghaus and Seth, 2007). These data cover almost complete (un-)employment and benefit histories for the period from 2005 to 2009.

More precisely, this study's aim is to estimate the causal effect of ME on the subsequent chances of regular employment for a group of persons who (1) receive UB II and (2) were unemployed at the beginning of the UB-II receipt compared to the outcome in which these persons did not take up ME. The latter outcome is obviously unobserved for the group of people who actually took up ME. Therefore, the potential outcome approach of causality (see, e.g., Heckman *et al.*, 1999; Roy, 1951; Rubin, 1974) is a natural point of departure for our analysis. In this approach, using standard notation, Y_1 is the potential outcome (in our case, probability of regular employment) if the person is treated (being marginally employed, $D = 1$), whereas Y_0 is the potential outcome if the person is not treated (not marginally employed, $D = 0$).

Recent literature notes the importance of the timing of treatment events and selection of an appropriate control group for the treated individuals (Abbring and van den Berg, 2003; Hujer and Thomsen, 2010; Sianesi, 2004; Stephan, 2008). For our application, taking up ME likely has a positive effect on regular employment because on the one hand, ME provides a form of occupational stabilization. On the other hand, there might be a negative lock-in effect of ME in the form of reduced search efforts for regular employment. The relative importance of these effects will arguably change with the time spent receiving benefits: The former effect might become more important the longer an individual is unemployed before taking up ME, whereas the lock-in effect might be relatively more important if a person takes up ME early after beginning to receive benefits. For these reasons, we follow the dynamic matching approach of Sianesi (2004), which allows the effect

of taking up ME on the probability of regular employment to change with the timing of the treatment. This approach requires that we discretize the time after entering benefit receipt. Let $U = \{u_1, u_2, \dots, u_{\max}\}$ be the set of elapsed durations after receiving UB II for the first time. Then, for each elapsed duration u , we estimate the dynamic average treatment effect on the treated, $ATT_{t,u}$, which is defined as the mean of the differences between the outcomes for persons in period $t > u$, who took up ME after an elapsed duration u of receiving UB II and the outcomes in the counterfactual situation, in which these persons would not have been treated at least until period u . Formally,

$$\begin{aligned} ATT_{t,u} &= E\left(Y_{t,u}^1 - Y_{t,u}^0 | D_u = 1, D_1 = \dots = D_{u-1} = 0\right) \\ &= E\left(Y_{t,u}^1 | D_u = 1, D_1 = \dots = D_{u-1} = 0\right) \\ &\quad - E\left(Y_{t,u}^0 | D_u = 1, D_1 = \dots = D_{u-1} = 0\right), \end{aligned} \quad [1]$$

with $u = u_1, u_2, \dots, u_{\max}$, where $Y_{t,u}^1$ is the potential outcome for time t of a person who took up ME after u , whereas $Y_{t,u}^0$ is the corresponding potential outcome for a person who did not take up ME at least up to time u . Thus, in the dynamic matching approach, there is no clear-cut distinction between treated and non-treated persons: everyone who is regarded as non-treated after a specific period u may be treated at a later period. In contrast, in a non-dynamic matching approach, the control group would instead only consist of persons who never take up ME in the treatment period. As noted in, e.g., Hujer and Thomsen (2010), this latter definition of the control group might bias the estimated treatment effect because it is conditioned on future outcomes. For example, if individuals are never observed to take up ME because they find regular employment before doing so, the treatment effect of taking up ME will be negatively biased.

Whereas the first term of the last expression in [1] is identified by the data, the second term must be estimated. In a non-experimental study such as this one, simply substituting the counterfactual outcome $E\left(Y_{t,u}^0 | D_u = 1, D_1 = \dots = D_{u-1} = 0\right)$ with the (observed) mean outcome of persons untreated until u , $E\left(Y_0 | D_1 = \dots = D_u = 0\right)$ will likely lead to selection bias, i.e.,

$$E\left(Y_{t,u}^1 | D_u = 1, D_1 = \dots = D_{u-1} = 0\right) - E\left(Y_{t,u}^0 | D_1 = \dots = D_u = 0\right) \neq 0, \quad [2]$$

because individual characteristics that determine the outcome will typically also determine the treatment decision. Thus, for an unbiased estimate of the ATT in non-experimental situations, identifying assumptions to solve the problem of self-selection must be invoked.

In recent years, propensity score matching has become the standard approach in the literature on programme evaluation. Intuitively, this approach involves matching each treated individual to ‘statistical twins’, i.e., non-treated individuals with similar observed characteristics X , such that differences in the outcomes of both groups can be attributed to the treatment. As suggested by Rosenbaum and Rubin (1983), we match on the propensity score $p(X) = P(D = 1 | X)$, i.e., the probability of being treated given X .

Following this approach, we apply dynamic propensity score matching to estimate the $ATT_{t,u}$ of taking up ME on the probability of regular employment.⁷ Propensity score matching identifies the $ATT_{t,u}$ if two conditions are satisfied: the CIA and the common support condition (see, e.g., Caliendo and Kopeinig, 2008). For the case of dynamic

matching, an adjusted dynamic version of the CIA (DCIA) can be stated as (Hujer and Thomsen, 2010):

$$Y_{t,u}^0 \coprod D_u | p(X_u), D_1 = \dots = D_{u-1} = 0, \quad [3]$$

where \coprod denotes independence. This means that conditional on the propensity score $p(X_u)$ and not being treated up to time u , in the absence of treatment at u the treated would experience the same outcome as individuals from the control group. The common support condition is stated as $P(D = 1 | X_u) < 1$, which implies that non-treated matches for the treated must be available.

Because the DCIA assumption is fundamentally untestable, we must credibly argue why that assumption is likely to be valid for our application. On the one hand, in a dynamic matching approach, the DCIA is generally more likely to hold than the CIA in a static matching approach (Sianesi, 2004). This is the case because the DCIA only must hold at the margin (taking up ME at u versus taking up ME later) and not, like the CIA, once and for all, i.e., taking up ME versus not taking up ME up to u_{max} . Sianesi (2004) also notes that the dynamic matching approach reduces heterogeneity compared to the static approach because the current unemployment duration is controlled for, and this unemployment duration can be considered to capture further unobserved differences among individuals. However, the DCIA only holds if we observe all covariates X that jointly influence the participation decision at time u and the outcome variable where taking up ME is postponed further ($Y_{t,u}^0$), conditional on not taking up ME up to time u . However, for our application, we believe that the DCIA is likely to hold for the following reasons: First, because we focus on labour market outcomes, it is particularly important to employ variables on an individual's employment history for matching (see, e.g., Lechner, 1999). Thus, we match on past labour market outcomes measured as the number of days spent in full-time work, part-time work, ME, apprenticeship, job search, and active labour market programmes (ALMP) 1 year and 5 years before entering UB II, along with the number of days spent in unemployment before entering UB II. Second, to control for business cycle effects, we additionally match on the quarter of first entry into UB II. Finally, we match on several sociodemographic variables, i.e., age at the time of entering UB II, nationality and qualification, and regional labour market conditions.

Furthermore, we restrict our estimation sample to persons who should only exhibit negligible differences in unobserved heterogeneity. Although this selection comes at the price of reducing (to an extent) the generalizability of our results, it increases the likelihood that the DCIA will hold in our analysis. We select the sample as follows. First, our sample consists of persons who received UB II for the first time between 2005 and 2006.⁸ We confine the sample to this entry cohort because this allows us to observe UB-II recipients for as long as possible. Second, we restrict the sample to UB-II recipients who are actually at risk of transitioning into regular employment. Alternative roles and duties within the household context might influence the desired allocation of time among employment, housework, and leisure time — particularly for women with children and care responsibilities (Becker, 1965; Blau *et al.*, 2001). Therefore, we only select childless singles at the time of entry into UB II who are actually looking for full-time employment. This selection is possible because our administrative data on UB-II receipt include household and job search information, which was unavailable before the introduction of the Hartz IV reform.

Third, to focus on the core group of the labour force, we only consider persons who were 25–55 years old at the time of receiving UB II for the first time. Finally, because there remain remarkable differences in the labour market conditions between East and West Germany and because our database is less reliable for East Germany because of local labour agencies' high level of reporting failures in the first years after the introduction of UB II, we restrict our analysis to West Germany.

Our selection of the estimation sample should substantially reduce differences in motivations and restrictions related to job search. The selection leads to a sample size of 6,506 men and 2,669 women entering UB II for the first time between 2005 and 2006.

After entering benefit receipt, we distinguish four states that individuals can enter: (i) searching (waiting), (ii) taking up ME, (iii) taking up regular employment, and (iv) entering an ALMP programme. After entering the benefit receipt, all individuals begin in state (i) and are at risk of entering any of the states (ii) to (iv) before the end of each period $u = u_1, u_2, \dots, u_{max}$. We regard the states from (ii) to (iv) as absorbing, i.e., we only consider the first spell of types (ii) to (iv) after entering benefit receipt and not the whole sequence of spells in the four possible states. Thus, in each period u , the group of the treated consists of all of the individuals who take up ME for the first time before the end of u , whereas their respective control group consists of all of the individuals who (1) did not take up ME in u and (2) did not enter states (iii) or (iv) before the end of u . We estimate the propensity scores for entering ME, $p(X_u)$, by a sequence of multinomial logit models for each period u . This approach can be considered equivalent to a discrete-time hazard-independent competing risk model (Voßemer and Schuck, 2016), with all of the estimated parameters allowed to be duration specific (Sianesi, 2004). We choose the maximum duration for transitioning to states (ii) to (iv) after entering benefit receipt, u_{max} as 9 months. Individuals who are still searching for a job after 9 months are treated as right censored. This choice results in 1,125 men and 627 women who first take up ME after entering benefit receipt and captures approximately 80 per cent of all individuals in the sample who first take up ME in the first 2 years after entering benefit receipt. Simultaneously, a u_{max} of 9 months enables us to follow all of the individuals in our sample for 36 months after each period u . More precisely, we estimate the $ATT_{t,u}$ for $t = 6, 12, \dots, 36$ months after first entering ME at $u = u_1, \dots, u_{max}$. Our sample size does not allow us to estimate the $ATT_{t,u}$ for $u = 1, 2, \dots, 9$ months because this choice leads to a very low number of treated individuals in the later months. Instead, we choose periods u_i , $i = 1, \dots, I$ of unequal length such that the share of the treated individuals is approximately equal in each period. As a robustness check, we choose both $I = 3$ and $I = 4$ periods.⁹

Given the validity of the DCIA and estimated propensity scores $p(X_u)$, the ATT can be consistently estimated by computing the mean of the difference of the outcomes for the group of treated individuals (or any subgroup thereof) and a (weighted) control group of non-treated individuals: $ATT_{t,u} = (1/N_{1,u}) \sum_{i=1}^{N_{1,u}} (Y_{1i,t} - \sum_{j=1}^{N_{0,u}} w_{ij,u} Y_{0j,t})$, where $N_{0,u}$ and $N_{1,u}$ are the number of observations in the control group and the treatment group in period u , respectively, and $w_{ij,u}$ are the weights for the outcomes of the j -th individual of the control group used to estimate the ATT for the i -th treated individual. The size of the weights depends on the matching algorithm used.

Asymptotically, all available matching algorithms — given the validity of the DCIA assumption — will lead to unbiased estimates of $ATT_{t,u}$. Moreover, there is no superior matching algorithm in finite samples.¹⁰ All of the available algorithms can instead be considered a trade-off between the bias and the variance of the estimated ATT (Caliendo and Kopeinig, 2008). In this application, we first employ kernel matching (KM)¹¹ with the

Epanechnikov kernel and a bandwidth of 0.02. We chose the bandwidth with the aim to approximately optimize the matching quality.¹²

In addition, following Hujer and Thomsen (2010), we use 1-to-1 nearest-neighbour (NN) matching without replacement. On the one hand, 1-to-1 NN matching without replacement constitutes the extreme opposite to KM in the trade-off between bias and variance of the estimated *ATTs*. Therefore, by employing both, KM and 1-to-1 NN matching without replacement, we check the robustness of the estimated *ATTs*. On the other hand, matching without replacement is required to apply the Mantel and Haenszel (1959) test (MH test) for the estimated *ATTs*' sensitivity to the presence of a possible hidden bias (Rosenbaum, 2002). As discussed above, the DCIA only holds if unobserved characteristics are irrelevant to both the treatment and the outcome. Although we argue that our data set is rich enough to describe the individual's labour market situation, we believe that it is nonetheless informative to test the sensitivity of the estimated *ATTs* with respect to possible unobserved selection ('hidden bias'). For example, if unobserved factors that increase the probability of being treated also increase the probability of a positive outcome, *ATTs* would be overestimated. Intuitively, the MH test calculates bounds for the 'amount' of hidden bias at which the inference about the treatment effect is altered. Thus, and crucially, the MH test cannot indicate whether a hidden bias is present; it merely shows how robust the estimated *ATTs* are with respect to a possible hidden bias.¹³

5. Results

Turning to the results of the analysis, we first show that our matching procedure can balance the treatment and control groups on our selected covariates.¹⁴ Table 1 displays several indicators for the quality of the KM procedure for men and women and the different time intervals (the two specifications with four and three periods displayed separately). First, the results show that in the case of four (3) treatment periods, there are an average of approximately 220 (290) treated men and 130 (170) treated women in each treatment period. Almost no individuals are lost because of the common support condition. The KM procedure can reduce bias in every single case to a sufficient degree, with a median bias of well below three after matching (Caliendo and Kopeinig, 2008), and a Pseudo- R^2 at almost zero. The results for 1-to-1 nearest-neighbour matching without replacement (NN) (Table 2) show less favourable — but still reasonably good — values. The median bias is reduced quite substantially, and the Pseudo- R^2 values are also reduced in all cases except for the two latest periods for women. In these two cases, the inspection of covariate-specific bias (not displayed here) shows that there is only one case in which one covariate is still significantly different between the treated and control groups after matching.

Given the good matching quality — particularly for KM — we can now turn to the results concerning the effects on regular employment. As described above, we measured the share of workers in regular employment every 6 months following their respective treatment period for up to 36 months. In Tables 3 and 4, KM results are reported not only for men and women but also for the 3- and 4-period specifications of u . We find a unanimous pattern in all groups: There are substantial treatment effects for the unemployed who enter ME in the later treatment periods, whereas there are almost no highly statistically significant effects for the unemployed taking up ME in the first searching period, i.e., within the first 30 or 40 days, respectively. For the last searching period, we find statistically highly

Table 1. Overview on matching quality — kernel matching

Searching period	No. of treated before	No. of non-treated before	Median bias before ^b	Median bias after ^b	Probit ps-R ² before ^c	Probit ps-R ² after ^c	CS ^d
Men							
<i>u</i> ^a : 4 Periods							
1: 0–30 days	225	5758	8.7	1.2	0.083	0.004	1
2: 31–70 days	206	4762	8.1	1.7	0.080	0.003	2
3: 71–150 days	236	3526	9	1.2	0.058	0.002	0
4: 151–270 days	209	2387	6.7	1.6	0.055	0.004	1
<i>u</i> ^a : 3 Periods							
1: 0–40 days	303	5435	8.3	1.2	0.076	0.003	1
2: 41–120 days	292	3928	9.6	2.2	0.070	0.003	0
3: 121–270 days	281	2387	5.6	2	0.061	0.005	1
Women							
<i>u</i> ^a : 4 Periods							
1: 0–30 days	130	2326	10.8	1.1	0.089	0.003	0
2: 31–70 days	143	1864	11.7	2.7	0.110	0.005	2
3: 71–150 days	126	1345	10.6	1.9	0.084	0.004	1
4: 151–270 days	113	890	10.6	1.6	0.061	0.006	0
<i>u</i> ^a : 3 Periods							
1: 0–40 days	182	2180	10	1.2	0.103	0.003	0
2: 41–120 days	170	1517	9.7	2.1	0.088	0.003	0
3: 121–270 days	160	890	11.9	2.2	0.060	0.008	0

^aPeriod *u* refers to the time (in days) at which the treatment starts in the individual UB-II spell.

^bMedian bias denotes the median of the standardised difference in per cent following Rosenbaum and Rubin (1985) before and after matching.

^cProbit ps-R² refers to the pseudo R² computed for the full sample (before) and the matched sample (after).

^dNumber of treated individuals lost after imposing the common support condition.

significant (1 per cent level) treatment effects for all outcome periods except one, for which the significance level is 5 per cent. This holds for both men and women, and the effects are of a substantial magnitude of approximately 10 to 20 percentage points. For example, for men in the 4-period case taking up ME after at least 150 days (5 months) spent in unemployment and UB-II receipt, the probability of being regularly employed 30 months after the treatment is 14 percentage points higher than for men who did not take up ME until the end of the at risk period (270 days after entering UB-II receipt). For women, the treatment effect is even larger, at 20 percentage points.

A very similar picture emerges in the NN case (Tables 5 and 6). Here, too, we find comparatively small and mostly insignificant treatment effects for the first searching periods of up to 30 or 40 days. In the later periods of unemployment durations — particularly after 150 days — there again are significant and large effects of between 10 and (in some cases) 25 percentage points.¹⁵ Compared to the KM results, the NN estimates do not reach the highest significance levels in as many cases for the group of women in the fourth period. This might be partially explained by the fact that we lose more treated observations for NN matching than for KM because of the common support assumption. Because the results obtained with two matching estimators at opposite extremes of the bias–variance trade-off (Caliendo and Kopeinig, 2008) are highly similar, it is safe to assume that the general pattern of effects holds, not at least because there are clear-cut explanations from a substantial perspective.

Table 2. Overview on matching quality — 1-to-1 nearest-neighbour matching

Searching period	No. of treated before	No. of non-treated before	Median bias before ^b	Median bias after ^b	Probit ps-R ² before ^c	Probit ps-R ² after ^c	CS ^d
Men							
<i>u</i> ^a : 4 Periods							
1: 0–30 days	225	5758	8.7	5.5	0.083	0.039	10
2: 31–70 days	206	4762	8.1	4.2	0.08	0.035	9
3: 71–150 days	236	3526	9	3.9	0.058	0.02	8
4: 151–270 days	209	2387	6.7	5.9	0.055	0.031	13
<i>u</i> ^a : 3 Periods							
1: 0–40 days	303	5435	8.3	3.4	0.076	0.016	14
2: 41–120 days	292	3928	9.6	3.2	0.070	0.022	12
3: 121–270 days	281	2387	5.6	4.2	0.061	0.028	21
Women							
<i>u</i> ^a : 4 Periods							
1: 0–30 days	130	2326	10.8	5.3	0.089	0.053	9
2: 31–70 days	143	1864	11.7	3.7	0.110	0.039	14
3: 71–150 days	126	1345	10.6	3.2	0.084	0.052	10
4: 151–270 days	113	890	10.6	7	0.061	0.062	12
<i>u</i> ^a : 3 Periods							
1: 0–40 days	182	2180	10.0	3.4	0.103	0.024	18
2: 41–120 days	170	1517	9.7	6.4	0.088	0.041	19
3: 121–270 days	160	890	11.9	7.6	0.060	0.062	16

^aPeriod *u* refers to the time (in days) at which the treatment starts in the individual UB-II spell.

^bMedian bias denotes the median of the standardized difference in per cent following Rosenbaum and Rubin (1985) before and after matching.

^cProbit ps-R² refers to the pseudo R² computed for the full sample (before) and the matched sample (after).

^dNumber of treated individuals lost after imposing the common support condition.

Finally, the general dynamic pattern of the treatment effects of our results is consistent with our theoretical discussion (see Section 3) of the ‘stepping stone’ function of ME because we argued that taking up ME later in the unemployment spell should likely have a more substantial effect on subsequent regular employment than does taking up ME after a relatively short period of unemployment.

6. Conclusion

In this study we analysed whether ME can improve the chances of regular employment for a special group of unemployed benefit recipients and applied a dynamic matching approach that is state of the art but has not yet been used for the ME analysis. The results presented here have shown the appropriateness of the dynamic evaluation approach when analysing the employment effects of ME on unemployed men and women. We find a clear-cut trend of substantial employment effects for unemployed men and women who remain unemployed after entering UB-II receipt for at least 4 or 5 months, whereas there are no unambiguous effects for individuals with shorter unemployment durations. These results are more or less in accordance with existing empirical evidence of the positive effects of ME for either unemployed individuals with higher unemployment duration (Caliendo *et al.*, 2012, 2016) or a sample of long-term unemployed UB-II recipients (Lehmer, 2012).

Table 3. Average treatment effects (*ATTs*) of ME on regular employment — kernel matching — 4 Periods

Searching period u^a	Treated ^b	Outcome period					
		$\Delta_{6,u}$	$\Delta_{12,u}$	$\Delta_{18,u}$	$\Delta_{24,u}$	$\Delta_{30,u}$	$\Delta_{36,u}$
Men							
1: 0–30 days	225	0.046	<i>0.060</i>	0.065	<i>0.067</i>	0.064	0.048
2: 31–70 days	206	0.048	<i>0.067</i>	0.034	0.033	0.066	0.065
3: 71–150 days	236	<i>0.119</i>	<i>0.083</i>	<i>0.141</i>	<i>0.111</i>	<i>0.100</i>	<i>0.105</i>
4: 151–270 days	209	<i>0.100</i>	<i>0.152</i>	<i>0.186</i>	<i>0.135</i>	<i>0.139</i>	<i>0.125</i>
Women							
1: 0–30 days	130	0.056	0.063	0.029	0.052	0.043	0.011
2: 31–70 days	143	0.060	<i>0.096</i>	<i>0.131</i>	<i>0.125</i>	<i>0.063</i>	0.028
3: 71–150 days	126	<i>0.099</i>	<i>0.149</i>	<i>0.106</i>	<i>0.145</i>	<i>0.159</i>	<i>0.169</i>
4: 151–270 days	113	0.111	<i>0.165</i>	<i>0.146</i>	<i>0.123</i>	<i>0.197</i>	<i>0.175</i>

Bold and italic values indicate significance at the 1 per cent level, bold values refer to the 5 per cent level, and italic values refer to the 10 per cent level. For the kernel estimates bootstrapped and bias-corrected standard errors are used.

^aPeriod u refers to the time (in days) at which the treatment starts in the individual UB-II spell.

^bTreated refers to the number of treated observations when using kernel matching. Common support is imposed by the minimum–maximum comparison.

Table 4. Average treatment effects (*ATTs*) of ME on regular employment — Kernel matching — 3 Periods

Searching period u^a	Treated ^b	Outcome period					
		$\Delta_{6,u}$	$\Delta_{12,u}$	$\Delta_{18,u}$	$\Delta_{24,u}$	$\Delta_{30,u}$	$\Delta_{36,u}$
Men							
1: 0–40 days	303	0.047	<i>0.064</i>	0.072	<i>0.057</i>	0.070	0.043
2: 41–120 days	292	<i>0.135</i>	<i>0.084</i>	<i>0.114</i>	<i>0.144</i>	<i>0.111</i>	<i>0.162</i>
3: 121–270 days	281	<i>0.090</i>	<i>0.123</i>	<i>0.156</i>	<i>0.111</i>	<i>0.132</i>	<i>0.131</i>
Women							
1: 0–40 days	182	0.066	<i>0.109</i>	0.074	<i>0.092</i>	<i>0.094</i>	0.024
2: 41–120 days	170	0.105	0.102	<i>0.127</i>	0.085	<i>0.110</i>	<i>0.128</i>
3: 121–270 days	160	<i>0.140</i>	<i>0.183</i>	<i>0.187</i>	<i>0.168</i>	<i>0.215</i>	<i>0.198</i>

Bold and italic values indicate significance at the 1 per cent level, bold values refer to the 5 per cent level, and italic values refer to the 10 per cent level. For the kernel estimates bootstrapped and bias-corrected standard errors are used.

^aPeriod u refers to the time (in days) at which the treatment starts in the individual UB-II spell.

^bTreated refers to the number of treated observations when using kernel matching. Common support is imposed by the minimum–maximum comparison.

The analysis used newly available data for UB-II recipients, thus enabling us to focus on a special group of unemployed men and women who should be equally likely to profit from ME. We restricted our analysis to the single and childless unemployed who are searching for full-time employment. Therefore, there should be no factors that could inhibit transitions to regular employment, such as a missing wish for extended employment, and restricting factors, such as family obligations. This construction of the sample made it

Table 5. Average treatment effects (*ATTs*) of ME on regular employment — 1-to-1 nearest-neighbour matching — 4 Periods

		Outcome period					
Searching period <i>u</i> ^a	Treated ^b	$\Delta_{6,u}$	$\Delta_{12,u}$	$\Delta_{18,u}$	$\Delta_{24,u}$	$\Delta_{30,u}$	$\Delta_{36,u}$
Men							
1: 0–30 days	215	0.042	0.019	0.037	0.033	0.047	0.051
2: 31–70 days	197	0.066	0.066	0.051	0.046	0.107	0.107
3: 71–150 days	228	<i>0.110</i>	0.101	<i>0.140</i>	<i>0.158</i>	<i>0.149</i>	<i>0.162</i>
4: 151–270 days	96	<i>0.107</i>	<i>0.148</i>	<i>0.204</i>	<i>0.143</i>	<i>0.173</i>	<i>0.128</i>
Women							
1: 0–30 days	121	0.066	0.124	0.050	0.099	0.091	0.041
2: 31–70 days	129	0.039	0.047	<i>0.116</i>	0.101	0.070	0.000
3: 71–150 days	116	0.060	0.147	<i>0.112</i>	0.155	0.103	0.103
4: 151–270 days	101	0.109	<i>0.178</i>	0.139	<i>0.109</i>	<i>0.228</i>	<i>0.228</i>

Bold and italic values indicate significance at the 1 per cent level, bold values refer to the 5 per cent level, and italic values refer to the 10 per cent level.

^aPeriod *u* refers to the time (in days) at which the treatment starts in the individual UB-II spell.

^bTreated refers to the number of treated observations when using kernel matching. Common support is imposed by the minimum–maximum comparison.

Table 6. Average Treatment Effects (*ATTs*) of ME on Regular Employment — 1-to-1 Nearest-Neighbour Matching — 3 Periods

Searching period <i>u</i> ^a	Treated ^b	Outcome period					
		Δ _{6,<i>u</i>}	Δ _{12,<i>u</i>}	Δ _{18,<i>u</i>}	Δ _{24,<i>u</i>}	Δ _{30,<i>u</i>}	Δ _{36,<i>u</i>}
Men							
1: 0–40 days	289	0.055	0.087	<i>0.114</i>	0.048	<i>0.073</i>	0.028
2: 41–120 days	280	<i>0.132</i>	0.039	0.064	0.121	<i>0.075</i>	0.146
3: 121–270 days	260	<i>0.104</i>	<i>0.131</i>	<i>0.150</i>	0.100	<i>0.131</i>	<i>0.146</i>
Women							
1: 0–40 days	164	0.030	0.079	0.043	0.043	0.104	0.067
2: 41–120 days	151	<i>0.093</i>	0.146	0.139	0.079	0.053	0.086
3: 121–270 days	144	<i>0.146</i>	<i>0.208</i>	<i>0.229</i>	<i>0.194</i>	<i>0.243</i>	<i>0.250</i>

Bold and italic values indicate significance at the 1 per cent level, bold values refer to the 5 per cent level, and italic values refer to the 10 per cent level.

^aPeriod *u* refers to the time (in days) at which the treatment starts in the individual UB-II spell.

^bTreated refers to the number of treated observations when using kernel matching. Common support is imposed by the minimum–maximum comparison.

possible to analyse women, who often have been excluded from the analysis because of the impossibility of controlling for household context and job search activities with administrative data. Our results show that single and childless women can profit from ME to a similar extent as men.

The generalizability of our results is somewhat limited by our sample selection. Whereas we find evidence that ME can improve chances for regular employment for single

unemployed individuals with longer unemployment durations, we cannot draw conclusions about individuals living in family households. Thus, a question for further research could be to compare results for the single unemployed with unemployed individuals in families who (in addition to job search) must balance work and family life as well. Nonetheless, our findings are highly relevant to the discussion of whether it is desirable for unemployed recipients of UB II to take up ME versus continue to search for a job in the hope of finding regular employment: singles constitute the clear majority among both the recipients of UB II and the marginally employed recipients of UB II.

Again, we want to stress that a causal interpretation of our results requires the strong (and untestable) CIA to hold. While we believe that our approach limits the possible bias introduced by unobserved heterogeneity as far as possible, readers should be cautious in interpreting our results as causal. Therefore, applying a different methodological approach that is able to explicitly model unobserved heterogeneity can be used to further examine the robustness of our results in future research.

Notes

¹ Marginal employment according to §8 Social Code, Book IV (SGB IV).

² This implies that marginally employed workers have no entitlement to social security benefits from health, unemployment, and pension insurances. Since 2013, ME has included liability for pension insurance contributions, although marginally employed workers can opt out of these contributions. In addition, the maximum labour income from ME was increased to 450 € per month in 2013. Otherwise, regulations for ME have remained fundamentally unchanged since 2003.

³ Until 2003, weekly working hours in ME were restricted to 15 hours. The introduction of the minimum wage of 8.50 € per hour in 2015 implicitly restricted the maximum number of weekly working hours in ME to approximately 12 hours.

⁴ If a worker takes up more than one secondary job, the combined income from the primary and secondary jobs is subject to taxes and social security contributions, even if the combined income from the secondary jobs does not exceed the maximum monthly labour income of 400 € (450 € since 2013) per month.

⁵ This line of argument is closely related to the discussion of the stepping stone or entrapment perspective on temporary employment (see, e.g., Gebel, 2013; Hagen, 2003; Korpi and Levin, 2001).

⁶ Thus, approximately 10 per cent of all UB-II recipients of working age (15–67 years) were marginally employed in recent years.

⁷ Regular employment is defined as either full- or part-time employment subject to social security contributions.

⁸ Persons who received UB II in the first month of its introduction, January 2005, are excluded from the sample to avoid transitions from the pre-UB-II system of social assistance.

⁹ For $I = 3$, we choose the following periods measured in days after entering UB II: 0–40, 41–120, and 121–270 days after entering benefit receipt. For the setting $I = 4$, we choose the periods 0–30, 31–70, 71–150, and 151–270 days after entering benefit receipt.

¹⁰ Huber *et al.* (2013) investigate the small sample properties of matching estimators in a Monte Carlo study. They find that a particular bias-adjusted radius matching estimator performs best overall. For our data, the performance of this estimator is comparable to the 1-to-1 nearest-neighbour matching without replacement, but not as good as the kernel matching estimator.

¹¹ For the matching, we use the programme ‘psmatch2’ (Leuven and Sianesi, 2003).

¹² Varying the bandwidth between 0.005 and 0.03 does not have a substantial effect on the matching quality and the estimated *ATT*s.

¹³ See Appendix B for a short description of the MH test.

¹⁴Selected results of the propensity score estimations using multinomial logit models are presented in Tables A1 and A2.

¹⁵The Mantel–Haenszel test (Table A3) shows that *ATTs* are quite robust with respect to a possible hidden bias. For the outcomes regular employment 30 and 36 months after treatment, Γ has to take on values of 1.3 or 1.5 for men to turn significant *ATTs* in the respective last periods into insignificant ones. For women, values of Γ are higher than 1.8. Therefore, we can assume the robustness of our effects even if there is a medium-sized hidden bias.

References

- Aakvik A. (2001) 'Bounding a Matching Estimator: The Case of a Norwegian Training Program', *Oxford Bulletin of Economics and Statistics* 63: 115–143.
- Abbring J. H. and van den Berg G. J. (2003) 'The Nonparametric Identification of Treatment Effects in Duration Models', *Econometrica* 71: 1491–1517.
- Albrecht J., Edin P., Sundström M. and Vroman S. (1999) 'Career Interruptions and Subsequent Earnings: A Reexamination Using Swedish Data', *Journal of Human Resources* 34: 294–311.
- Bäcker G. and Neuffer S. (2012) 'Von der Sonderregelung zur Beschäftigungsnorm: Minijobs im deutschen Sozialstaat', *WSI Mitteilungen* 2012/01: 13–21.
- Becker G. S. (1965) 'A Theory of the Allocation of Time', *The Economic Journal* 75: 493–517.
- Becker G. S. (1975) *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, New York: National Bureau of Economic Research with Columbia University Press.
- Becker S. O. and Caliendo M. (2007) 'Sensitivity Analysis for Average Treatment Effects', *The Stata Journal* 7: 71–83.
- Blau F., Ferber M. A. and Winkler A. E. (2001) *The Economics of Women, Men, and Work*. Englewood Cliffs: Prentice Hall.
- Bruckmeier K. and Wiemers J. (2011) 'A New Targeting? - A New Take-Up? Non-Take-Up of Social Assistance in Germany after Social Policy Reforms', IAB Discussion Paper 10/2011.
- Caliendo M. and Kopeinig S. (2008) 'Some Practical Guidance for the Implementation of Propensity Score Matching', *Journal of Economic Surveys* 22: 31–72.
- Caliendo M. and Wrohlich K. (2010) 'Evaluating the German 'Mini-Job' Reform Using a Natural Experiment', *Applied Economics* 42: 2475–2489.
- Caliendo M., Künn S. and Uhlendorff A. (2012) 'Marginal Employment, Unemployment Duration and Job Match Quality', IZA Discussion Paper 6499.
- Caliendo M., Künn S. and Uhlendorff A. (2016) 'Earnings Exemptions for Unemployed Workers: The Relationship between Marginal Employment', *Unemployment Duration and Job Quality*, *Labour Economics* 42: 177–193.
- Eichhorst W., Grienberger-Zingerle M. and Konle-Seidl R. (2010) 'Activating Labor Market and Social Policies in Germany: From Status Protection to Basic Income Support', *German Policy Studies* 6: 65–106.
- Freier R. and Steiner V. (2007) 'Marginal Employment': Stepping Stone or Dead End? Evaluating the German Experience', IZA Discussion Paper 3175.
- Freier R. and Steiner V. (2008) 'Marginal Employment': Stepping Stone or Dead End? Evaluating the German Experience', *Zeitschrift für Arbeitsmarktforschung* 2–3: 223–243.
- Gebel M. (2013) 'Is a Temporary Job Better Than Unemployment? A Cross-country Comparison Based on British, German, and Swiss Panel Data', *Schmollers Jahrbuch* 133: 143–156.
- Granovetter M. S. (1973) 'The Strength of Weak Ties', *American Journal of Sociology* 78: 1360–1380.
- Gregory M. and Jukes R. (2001) 'Unemployment and Subsequent Earnings: Estimating Scarring Among British Men 1984–94', *The Economic Journal* 111: 607–625.

- Hagen T. (2003) 'Do fixed-term contracts increase the long-term employment opportunities of the unemployed?', ZEW Discussion Paper No. 03–49.
- Heckman J. J., LaLonde R. J. and Smith J. A. (1999) 'The economics and econometrics of active labor market programs' in Ashenfelter O. and Card D. (eds.) *Handbook of Labor Economics*. Amsterdam: Elsevier: 1865–2097.
- Hipp L., Bernhardt J. and Allmendinger J. (2015) 'Institutions and the Prevalence of Nonstandard Employment', *Socio-Economic Review* 13: 351–377.
- Huber M., Lechner M. and Wunsch C. (2013) 'The Performance of Estimators Based on the Propensity Score', *Journal of Econometrics* 175: 1–21.
- Hujer R. and Thomsen S. (2010) 'How do the Employment Effects of Job Creation Schemes Differ with Respect to the Foregoing Unemployment Duration?', *Labor Economics* 17: 38–51.
- Jacobebbinghaus P. and Seth S. (2007) 'The German Integrated Employment Biographies Sample IEBS', *Schmollers Jahrbuch* 127: 335–342.
- Jacobi L. and Kluge J. (2007) 'Before and After the Hartz Reforms: The Performance of Active Labour Market Policy in Germany', *Zeitschrift für Arbeitsmarktforschung* 1: 45–64.
- Jahoda M. (1982) *Employment and Unemployment: A Social-psychological Analysis*. Cambridge: Cambridge University Press.
- Köhler M. (2015) *Benutzerhandbuch ASU. Arbeitsuchendenhistorik. Version 06.08.00-201504*, Nürnberg.
- Körner T., Meinken H. and Puch K. (2013) 'Wer sind die ausschließlich geringfügig Beschäftigten? Eine Analyse nach sozialer Lebenslage', *Wirtschaft und Statistik, Januar* 2013: 42–62.
- Korpi T. and Levin H. (2001) 'Precarious Footing: Temporary Employment as a Stepping Stone Out of Unemployment in Sweden', *Work, Employment and Society* 15: 127–148.
- Kvasnicka M. (2009) 'Does Temporary Help Work Provide a Stepping Stone to Regular Employment?' in Autor D. (ed.) *Studies of Labor Market Intermediation*. Chicago: The University of Chicago Press: 335–372.
- Lechner M. (1999) 'Nonparametric Bounds on Employment and Income Effects of Continuous Vocational Training in East Germany', *The Econometrics Journal* 2: 1–28.
- Lehmer F. (2012) 'Dient die Arbeitnehmerüberlassung für Langzeitarbeitslose als Brücke in nachhaltige Beschäftigung?', *Sozialer Fortschritt* 61: 190–197.
- Leuven E. and Sianesi B. (2003) *PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing*. version 4.0.11; available online at <https://ideas.repec.org/c/boc/bocode/s432001.html> (accessed 26 April 2017).
- Mantel N. and Haenszel W. (1959) 'Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease', *Journal of the National Cancer Institute* 22: 719–748.
- Mincer J. (1974) *Schooling, Experience and Earnings*. New York: National Bureau of Economic Research with Columbia University Press.
- Paul K. I. and Batinic B. (2010) 'The Need for Work: Jahoda's Manifest and Latent Functions of Employment in a Representative Sample of the German Population', *Journal of Organizational Behavior* 31: 45–64.
- Rosenbaum P. R. (2002) *Observational Studies*. New York: Springer.
- Rosenbaum P. R. and Rubin D. B. (1983) 'The Central Role of the Propensity Score in Observational Studies for Causal Effects', *Biometrika* 70: 41–50.
- Rosenbaum P. R. and Rubin D. B. (1985) 'Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score', *The American Statistician* 39: 33–38.
- Roy A. D. (1951) 'Some Thoughts on the Distribution of Earnings', *Oxford Economic Papers* 3: 135–145.
- Rubin D. B. (1974) 'Estimating Causal Effects to Treatments in Randomised and Nonrandomised Studies', *Journal of Educational Psychology* 66: 688–701.

- Rudolph H., Graf T., Koller L. and Lietzmann T. (2013) 'Anhang B2: Das Administrative Panel (AdminP) - Forschungsdaten mit Haushaltskontext zum SGB II' in Dietz M., Kupka P. and Ramos Lobato P. (eds.) *Acht Jahre Grundsicherung für Arbeitsuchende. Strukturen – Prozesse – Wirkungen*, Bielefeld: Bertelsmann: 351–357.
- RWI [Rheinisch-Westfälisches Institut für Wirtschaftsforschung] (2012) *Studie zur Analyse der geringfügigen Beschäftigungsverhältnisse*, Essen: RWI [Rheinisch-Westfälisches Institut für Wirtschaftsforschung].
- Sianesi B. (2004) 'An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s', *The Review of Economics and Statistics* 86: 133–155.
- Statistik der Bundesagentur für Arbeit (2011) *Analyse der Grundsicherung für Arbeitsuchende*, Nürnberg: Statistik der Bundesagentur für Arbeit.
- Stegmaier J., Gundert S., Tesching K. and Theuer S. (2015) 'Bezahlter Urlaub und Lohnfortzahlung im Krankheitsfall: In der Praxis besteht Nachholbedarf bei Minijobbern', *IAB-Kurzbericht*, 18/2015.
- Stephan G. (2008) 'The Effects of Active Labor Market Programs in Germany. An Investigation Using Different Definitions of Non-treatment', *Jahrbücher für Nationalökonomie und Statistik* 228: 586–611.
- Voßemer J. and Schuck B. (2016) 'Better Overeducated than Unemployed? The Short- and Long-Term Effects of an Overeducated Labour Market Re-entry', *European Sociological Review* 32: 251–265.

Appendix A Tables

Table A1. Discrete-time hazard competing risk duration model — Men — 4-Period Model — u : 0–30 days

	Marginal employment		Regular employment		ALMP ^a	
	Coef.	$P> z $	Coef.	$P> z $	Coef.	$P> z $
Time of entry (Ref.: 1st quarter 2005)						
2nd quarter 2005	−0.116	0.633	0.485	0.051	0.256	0.221
3rd quarter 2005	−0.027	0.914	−0.039	0.890	0.067	0.762
4th quarter 2005	0.062	0.806	0.191	0.501	0.298	0.201
1st quarter 2006	0.048	0.845	0.076	0.786	−0.241	0.347
2nd quarter 2006	0.299	0.272	0.614	0.029	−0.010	0.972
3rd quarter 2006	−0.170	0.605	0.120	0.717	0.336	0.224
4th quarter 2006	−0.135	0.689	−0.010	0.976	0.480	0.083
Qualification (Ref.: high)						
Low	0.692	0.025	−0.535	0.102	0.067	0.791
Medium	0.495	0.070	0.123	0.599	0.038	0.853
Age	−0.014	0.154	−0.026	0.011	−0.008	0.371
Nationality (Ref.: German)						
Not German	0.346	0.040	0.030	0.878	−0.164	0.358
Past (un-) employment experience (days)						
Duration of unemployment prior to UB-II entry	0.000	0.963	−0.001	0.388	0.002	0.013
Duration of unemployment prior to UB-II entry squared	0.000	0.773	0.000	0.607	0.000	0.022

Table A1. Continued

	Marginal employment		Regular employment		ALMP ^a	
	Coef.	$P> z $	Coef.	$P> z $	Coef.	$P> z $
Duration of full-time employment last year	0.003	0.018	0.002	0.135	0.002	0.165
Duration of part-time employment last year	0.009	0.000	0.005	0.004	−0.001	0.830
Duration of marginal employment last year	0.008	0.000	−0.001	0.730	0.002	0.343
Duration of training participation last year	0.004	0.023	0.008	0.000	0.005	0.002
Duration of job search w/o unemployment last year	0.002	0.591	0.006	0.018	0.004	0.135
Duration of job search with unemployment last year	0.001	0.359	0.001	0.308	0.003	0.021
Duration of ALMP participation last year	−0.001	0.784	0.000	0.812	0.011	0.000
Duration in other states last year	0.001	0.719	−0.003	0.455	−0.006	0.228
Duration of full-time employment last 5 years	0.000	0.108	0.000	0.045	0.000	0.862
Duration of part-time employment last 5 years	0.000	0.516	0.000	0.867	0.000	0.500
Duration of marginal employment last 5 years	0.001	0.061	0.000	0.415	0.001	0.110
Duration of Training participation last 5 years	0.000	0.963	−0.001	0.078	0.000	0.883
Duration of job search w/o unemployment last 5 years	−0.003	0.189	−0.002	0.210	0.000	0.759
Duration of job search with unemployment last 5 years	0.000	0.712	0.000	0.932	0.000	0.084
Duration of ALMP participation last 5 years	0.000	0.447	0.000	0.808	0.000	0.598
Duration in other states last 5 years	0.000	0.961	0.001	0.463	0.001	0.600
Regional unemployment rate	−0.019	0.308	−0.063	0.001	0.004	0.789
Constant	−4.242	0.000	−2.604	0.000	−4.628	0.000
Pseudo R-Square	0.1031					
Log Likelihood	−2786.7223					
<i>N</i>	6,498					

^aActive labour market policies.

Table A2. Discrete-time hazard competing risk duration model — Women — 4-Period Model — *u*: 0–30 days

	Marginal employment		Regular employment		ALMP ^a	
	Coef.	<i>P</i> > <i>z</i>	Coef.	<i>P</i> > <i>z</i>	Coef.	<i>P</i> > <i>z</i>
Time of entry (Ref.: 1st quarter 2005)						
2nd quarter 2005	0.087	0.798	0.782	0.111	0.825	0.055
3rd quarter 2005	−0.222	0.545	0.848	0.080	0.664	0.133
4th quarter 2005	0.154	0.660	0.796	0.114	0.004	0.994
1st quarter 2006	−0.119	0.740	0.617	0.224	0.172	0.724
2nd quarter 2006	0.361	0.326	1.293	0.009	1.139	0.016
3rd quarter 2006	0.277	0.479	1.020	0.054	1.034	0.034
4th quarter 2006	−0.327	0.517	0.474	0.433	1.137	0.024
Qualification (Ref.: high)						
Low	0.666	0.044	−0.812	0.065	−0.913	0.056
Medium	0.290	0.280	−0.386	0.129	−0.272	0.305
Age	0.013	0.236	−0.021	0.152	0.000	0.997
Nationality (Ref.: German)						
Not German	−0.188	0.515	−0.287	0.434	0.005	0.990
Past (un-) employment experience (days)						
Duration of unemployment prior to UB-II entry	−0.003	0.228	−0.001	0.715	0.001	0.721
Duration of unemployment prior to UB-II entry squared	0.000	0.746	0.000	0.960	0.000	0.913
Duration of full-time employment last year	0.002	0.274	0.001	0.541	0.002	0.485
Duration of part-time employment last year	0.000	0.906	0.003	0.133	−0.005	0.219
Duration of marginal employment last year	0.008	0.000	0.001	0.596	0.002	0.397
Duration of training participation last year	0.004	0.099	0.005	0.010	0.003	0.347
Duration of job search w/o unemployment last year	−0.003	0.323	−0.001	0.859	0.014	0.021
Duration of job search with unemployment last year	0.001	0.447	−0.001	0.694	0.003	0.273
Duration of ALMP participation last year	0.001	0.706	−0.002	0.552	0.012	0.000
Duration in other states last year	0.006	0.408	−0.004	0.570	−0.001	0.820
Duration of full-time employment last 5 years	0.000	0.753	0.000	0.162	0.001	0.250
Duration of part-time employment last 5 years	0.000	0.179	0.000	0.928	0.001	0.009
Duration of marginal employment last 5 years	−0.001	0.101	0.000	0.501	0.001	0.147
Duration of training participation last 5 years	−0.001	0.069	0.000	0.335	0.001	0.171
Duration of job search w/o unemployment last 5 years	0.002	0.116	0.002	0.103	−0.008	0.109
Duration of job search with unemployment last 5 years	0.000	0.819	0.001	0.188	0.001	0.086

Table A2. Continued

	Marginal employment		Regular employment		ALMP ^a	
	Coef.	$P> z $	Coef.	$P> z $	Coef.	$P> z $
Duration of ALMP participation last 5 years	0.001	0.090	0.000	0.907	0.000	0.938
Duration in other states last 5 years	−0.005	0.353	0.002	0.333	0.004	0.010
Regional unemployment rate	−0.039	0.123	−0.063	0.031	−0.063	0.028
Constant	−3.350	0.000	−2.769	0.001	−5.092	0.000
Pseudo R-Square	0.1173					
Log Likelihood	−1214.7598					
<i>N</i>	2,662					

^aActive labour market policies.**Table A3.** Sensitivity of the estimates to possible hidden bias — Mantel–Haenszel Test

		$\Delta_{30,u}$				$\Delta_{36,u}$			
		$Q_{MH} \Gamma = 1$	Value ^a of Γ	Bounds for Q_{MH}		$Q_{MH} \Gamma=1$	Value ^a of Γ	Bounds for Q_{MH}	
				Q_{MH}^+	Q_{MH}^-			Q_{MH}^+	Q_{MH}^-
Men									
4 Periods									
1: 0–30 days	n.s.					n.s.			
2: 31–70 days	2.452	1.25	1.368	3.551	2.461	1.30	1.382	3.556	
3: 71–150 days	3.840	1.60	1.422	6.326	4.058	1.70	1.353	6.858	
4: 151–270 days	3.497	1.60	1.311	5.754	2.557	1.30	1.337	3.798	
3 Periods									
1: 0–40 days	n.s.				n.s.				
2: 41–120 days	2.975	1.30	1.441	4.527	2.206	1.15	1.386	3.032	
3: 121–270 days	3.055	1.35	1.449	4.691	3.456	1.50	1.310	5.657	
Women									
4 Periods									
1: 0–30 days	n.s.				n.s.				
2: 31–70 days	n.s.				n.s.				
3: 71–150 days	1.321	1.00			n.s.				
4: 151–270 days	3.305	1.85	1.315	5.422	3.256	1.80	1.320	5.301	
3 Periods									
1: 0–40 days	n.s.				1.998	1.15	1.374	2.631	
2: 41–120 days	n.s.				1.387	1.00			
3: 121–270 days	4.211	2.05	1.301	7.326	4.305	2.15	1.282	7.543	

^a $\Gamma = 1$ denotes the case of no hidden bias. The higher Γ can be increased without altering inference about Q_{MH}^+ or Q_{MH}^- , the more robust is the estimated *ATT* with respect to hidden bias. n.s. denotes insignificant treatment effects.

Appendix B Mantel–Haenszel (MH) Test

Formally, the treatment probability of a treated individual is given by the propensity score $p_i = P(X_i, u_i) = P(D_i = 1 | X_i, u_i) = F(\beta X_i + \gamma v_i)$, where v_i denotes an unobserved variable and γ measures the impact on that variable on the treatment probability. If there is no hidden bias, $\gamma = 0$ and the treatment will be solely determined by the observed variables X_i . Given a matched individual j with $X_i = X_j$ and assuming that $F(\cdot)$ is the logistic distribution, the odds ratio of treatment is given by

$$\frac{p_i(1-p_j)}{p_j(1-p_i)} = \frac{\exp(\beta X_i + \gamma v_i)}{\exp(\beta X_j + \gamma v_j)} = \exp(\gamma[v_i - v_j]).$$

Following Aakvik (2001) in assuming that the unobserved influence takes the form of a dummy, $v_i, v_j \in \{0, 1\}$, Rosenbaum (2002) shows that the odds ratio is bounded by

$$\frac{1}{\exp(\gamma)} \leq \frac{p_i(1-p_j)}{p_j(1-p_i)} \leq \exp(\gamma).$$

In the case of no hidden bias, $\gamma = 0$, and both individuals have the same probability of being treated. However, if $\gamma \neq 0$, the odds of being treated could differ by a factor of at most $\exp(\gamma) \equiv \Gamma$. In this sense, Γ is a measure of the departure from the assumption of no hidden bias. Aakvik (2001) shows that the MH test can be used to test the null hypothesis of no treatment effect. The MH test statistic asymptotically follows a standard normal distribution and is given by

$$Q_{MH} = \frac{|y_1 - E(y_1)| - 0.5}{\sqrt{Var(y_1)}},$$

with $E(y_1) = (n_1 y)/n$ and $Var(y_1) = [n_1 n_0 y(n-y)]/[n^2(n-1)]$, where n_1 (n_0) denotes the number of treated (matched non-treated), and y_1 (y_0) is the number of treated (matched non-treated) individuals with a successful outcome. Finally, $n = n_0 + n_1$ and $y = y_0 + y_1$. Rosenbaum (2002) shows that the MH test statistic can be bounded by two statistics Q_{MH}^+ and Q_{MH}^- , given by

$$Q_{MH}^+ = \frac{|y_1 - \tilde{E}^+| - 0.5}{\sqrt{Var(\tilde{E}^+)}} \leq Q_{MH} \leq Q_{MH}^- = \frac{|y_1 - \tilde{E}^-| - 0.5}{\sqrt{Var(\tilde{E}^-)}},$$

where \tilde{E} and $Var(\tilde{E})$ are the large-sample approximation to the expectation and variance of the number of treated persons with a successful outcome when v is binary and for a given γ . The statistics coincide for $\Gamma = 1$ and move apart for increasing Γ . If, for example, the estimated *ATT* is positive and significant under the assumption of no hidden bias ($\Gamma = 1$), the robustness of the *ATT* can be tested by increasing Γ (starting with 1) in small increments and finding the Γ at which the statistic Q_{MH}^+ becomes insignificant. On the other hand, in case of a significantly negative *ATT*, Γ is increased until the statistic Q_{MH}^- becomes insignificant. The higher Γ can be increased without altering inference about Q_{MH}^+ or Q_{MH}^- , the more robust is the estimated *ATT* with respect to hidden bias. We apply the MH test with the programme ‘mhbounds’ by Becker and Caliendo (2007).