Minimum Wage and Employment: Escaping the Parametric Straitjacket[*]

Stefano Cabras[†], Jan Fidrmuc[†] and J.D. Tena[‡]

April 2017

Abstract

Parametric regression models are often not flexible enough to capture the true relationships as they tend to rely on arbitrary identification assumptions. Using the UK Labor Force Survey, we estimate the causal effect of national minimum wage (NMW) increases on the probability of job entry and job exit by means of a non-parametric Bayesian modelling approach known as Bayesian Additive Regression Trees (BART). The application of this methodology has the important advantage that it does not require ad-hoc assumptions about model fitting, number of covariates or how they interact. We find that the NMW exerts a positive and significant impact on both the probability of job entry and job exit. Although the magnitude of the effect on job entry is higher, the overall effect of NMW is ambiguous as there are many more employed workers. The causal effect of NMW is found to be higher for young workers and in periods of high unemployment. On the other hand, no significant interactions were found with gender and qualifications.

Keywords: BART; causal inference; regression approach; matching regression.

JEL Codes: C23, C11, C14, J3, J4.

---

[*] We received helpful comments and suggestions from Brendan McCabe, Andy Tremayne and David Forrest, as well as seminar/conference participants at the University of Liverpool and "The German Minimum Wage – First Evidence and Experiences from Other Countries" conference at the Institute for Employment Research (IAB) in Nuremberg. Comments and suggestions from two anonymous referees are gratefully acknowledged.

[†] Departamento de Estadística, Universidad Carlos III and Dipartimento di Matematica e Informatica, Università di Cagliari. Contact information: C/Madrid 126. 28903 Getafe (Madrid), Spain, stefano.cabras@uc3m.es. Phone: +34 916249314.

† Corresponding Author. Department of Economics and Finance and Centre for Economic Development and Institutions (CEDI), Brunel University; Institute of Economic Studies, Charles University; and CESifo Munich. Contact information: Department of Economics and Finance, Brunel University, Uxbridge, UB8 3PH, United Kingdom. Jan.Fidrmuc@brunel.ac.uk or jan@fidrmuc.net. http://www.fidrmuc.net/. Phone: +44-1895-266-528.

‡ Management School, University of Liverpool; Dipartimento di Scienze Economiche e aziendali, Università di Sassari; CRENoS; and Instituto Flores de Lemus Contact information: Chatham Street, Liverpool L69 7ZH, UK, jtena@liv.ac.uk. Phone: +34 916241259 .

1 Introduction

The most characteristic feature of the literature on the causal impact of the minimum wage on employment is the general lack of consensus. Neumark and Washer (2007) compile an extensive survey of previous research and conclude that the minimum wage exerts an adverse impact on employment of low-skilled workers and a non-significant impact on total employment. However, other surveys on this issue, a meta-analysis by Card and Krueger (1995) and the subsequent contributions by Doucouliagos and Stanley (2008) and De Linde et al. (2014) find that there is a wide range of results in the previous research, and that once the publication selection bias is accounted for the mean estimate is consistent with a non-significant impact of the minimum wage on employment.

A possible reason for the wide range of findings is the fact that the results hinge dramatically on ad hoc assumptions about the parametric specification of the empirical model and on the definition of the control group in the analysis. This is corroborated in the insightful and interesting discussion in a series of papers by Allegretto et al. (2011, 2013), Dube et al. (2010), and Neumark et al. (2014) in a state-level panel analysis for the US. Dube et al. (2010) and Allegretto et al. (2011) suggest that it is essential to control for spatial heterogeneity in order to estimate the impact of the minimum wage in a panel data setting. In particular, they propose to include two types of local controls consisting of: (1) jurisdiction-specific linear time trends; and (2) interactions between time dummy variables for sets of neighboring states or neighboring counties so they could be used as controls to determine the impact of the minimum wages. Subsequently, Neumark et al. (2014) and Sabia et al. (2013) criticize these measures on the grounds that there are other non-linear ways of controlling for unobserved trends and that this approach excludes other potential controls apart from those for the neighboring regions. Crucially, the parametric form of the model appears to be the critical determinant of whether a significant or insignificant impact of minimum wage on employment is obtained. Hirsch et al. (2015), in turn, argue that the lack of a significant effect can be driven by alternative channels of adjustment such as changes in prices, profits, performance standards, and wage compression.

Another potential problem of the minimum-wage literature mentioned above is the fact that many studies analyze this issue using aggregate data.[1] Aggregation might mask the real effect of minimum wage at the individual level. Moreover, the analyses based on aggregate data could be affected by endogeneity as mimimum wage movements could be caused by regional or national macroeconomic variables (Baskaya and Rubinstein, 2011; Sabia, 2014). While policy variables can be endogenous to aggregate employment indicators, they are clearly exogenous with respect to specific individuals and their outcomes.

In this paper, we use the UK Labor Force Survey to estimate the causal impact of the UK national minimum wage (NMW) on employment using a non-parametric Bayesian modelling approach known as the Bayesian Additive Regression Trees (BART henceforth) that was originally developed by Chipman et al. (2010) and applied to the analysis of causal inference by Hill (2011), Sparapani et al. (2016), Tan et al. (2016) and others. This procedure shares some similarities with standard matching estimation strategies (see for example Abadie and Imbens, 2006), as it compares unemployment-to-employment and employment-to-unemployment transitions of individuals affected by the NMW increase with similar individuals who are unaffected by the increase but are sufficiently similar to the treatment group. The BART procedure has important advantages over other more traditional parametric specifications. Among them, it does not require any type of hypotheses or priors over the covariates to be included in the model, it can consider a large number of regressors, and it can estimate any type of interactive effects between the treatment variable and any other variable. Thus, under the BART model, the definition of the closest untreated individual for each treated individual and the interactions between the different clusters of individuals and time or and any other relevant covariate is not constrained to follow any specific (and potentially ad-hoc) parametric function. Furthermore, and more importantly, the parametric function need not be specified a priori.

[1] See for example Lee et al (1990) for a discussion on aggregation bias.

Our paper is closely related to at least two previous works that estimate the impact of the UK NMW on employment at the individual level using micro-data: Stewart (2004) and Dickens and Draca (2005). Stewart (2004) analyzes how the introduction of the UK NMW in 1999 and its subsequent changes in 2000 and 2001 affected the employment-to-employment transition. Dickens and Draca (2005) follow a similar approach for the NMW increases in October 2003 but they extend the analysis to consider the separate effect of the NMW on job entry and job exit decisions. Both study the impact of the NMW by applying the difference-in-difference technique to the UK Labor Force Survey data, and find that the NMW does not have a significant adverse effect on employment. Unlike these papers, we do not consider a specific year's increase in the minimum wage but take into account all NMW changes since its introduction in 1999. Finally, our approach allows us to identify the interactions of the NMW effect with other relevant variables such as gender, age, qualifications and business cycle without the necessity of proposing a parametric specification.

The contribution of our paper is twofold. First, we shed some new light on the relationship between the minimum wage and employment. In particular, we find that the NMW exerts a positive and significant impact on both the probability job entry and job exit. Although the magnitude of the effect on job entry is larger, the overall effect of NMW is ambiguous as there are many more employed than unemployed workers. This could explain the insignificant effect found in the previous work based on aggregate macroeconomic estimations. We find also that the effect is stronger for younger workers and in high unemployment periods. On the other hand, gender and qualifications play little role in shaping the minimum wage effect.

Second, we demonstrate the applicability of the BART approach to analyses of economic outcomes without imposing a specific parametric form a priori. While we chose the minimum wage effect on employment, this method could be applied to a broad range of other contexts equally well.

In the next section, we present the data used. Sections 3 and 4 discuss methodological approaches used for analyzing the labor-market impact of the minimum wage and explain the main features of the BART model, respectively. Empirical results are shown and discussed in Section 5. The final section summarizes our findings and offers some conclusions.

2 Data

Our analysis is based on the UK Labor Force Survey (LFS). The LFS is a quarterly nationally-representative survey of households across the UK. Each quarter, approximately 60 thousand households and over 100 thousand individuals aged 16 and above are surveyed. Each household is retained in the survey for five consecutive quarters, with one-fifth of households replaced in each wave. The survey contains detailed demographic and socio-economic information on the respondents, including, importantly, their labor-market outcomes. Since the NMW was introduced in April 1999, we use all quarterly datasets available from April-June 1999 to October-December 2011, pooling all available LFS waves during this period. In order to have a sufficient number of observations, we include all individuals aged between 16 and 40.

The UK NMW features three different age-dependent rates: the 16-17 years old rate, the youth rate (applying to those aged 18-21[2]), and the adult rate.[3] Historically, the youth rate has remained some 35% higher than the 16-17 rate while the adult rate has exceeded the youth rate by around 20%. The LFS reports the date of birth of every respondent and also the date the survey was carried out. By comparing these two dates, we can determine the precise age of each respondent on the day of the survey.[4] We therefore know whether a particular individual is below or above the age threshold at which they become eligible for a different (higher) NMW rate.

[2] The upper limit for the youth rate has been lowered to 20 from October 2010. Where relevant, our analysis takes this change into account.
[3] A fourth rate, for apprentice workers, was introduced in October 2010 (we do not consider those subject to this rate in our analysis). No minimum wage applies to those who belong to one of the few exemptions such as members of the armed forces, volunteers, students on work placements, workers living in the employers' households, and (until 2010) apprentices.
[4] The precise date of birth is not available in the publicly released LFS datasets. We are grateful to the Office for National Statistics for making the restricted release of the LFS available to us.

3 Methodological Considerations

We analyze the effect of the NMW increases on employment by going beyond standard regression and matching estimation methodologies traditionally used for this purpose. Regardless of the methodology, the analysis involves comparing the changes in labor-market outcomes (such as employment) after a NMW change for the treatment and control groups.[5] Consider the impact of NMW on the probability of job loss. The treatment group comprises workers whose wages have to go up in the wake of an annual NMW increase because the new NMW rate is higher than their current wage. The wages of those in the control group should be close to but just above the new rate so as not to have to change.

More specifically, the treatment group can be defined as the individuals whose wages meet the following condition:

$$\mathrm{nmw_t} < \mathrm{w_{it}} < \mathrm{nmw_{t+1}} \tag{1}$$

where $\mathrm{nmw_t}$ is the (age-dependent) NMW rate in effect at time t while $\mathrm{w_{it}}$ is the worker *i*'s wage. The control group is defined as the workers whose wage before the increase is greater than the new NMW rate but lower than some upper bound to ensure that we only consider workers earning just above the minimum wage (who, therefore, are likely to display similar characteristics as those earning the minimum wage). If we set the upper bound as a fraction *c* above the new rate. The control group thus comprises workers meeting the following condition:

$$\mathrm{nmw_{t+1}} \leq \mathrm{w_{it}} < \mathrm{nmw_{t+1}} * (1 + \mathrm{c}) \tag{2}$$

We can then estimate the following equation

$$P(e_{t+1} = 0 | e_t = 1) = \Phi(\alpha * D_i + \gamma * X_{it}) \tag{3}$$

[5] Note that our approach is similar in spirit to the difference-in-difference approach in Stewart (2004) and Dickens and Draca (2005) who compare the average change in the employment status before and after the introduction of a very specific minimum wage policy. Of course, as we show in Table 1, the vector X_{it} incorporates time-invariant characteristics. Given that we consider a whole sample of minimum wage changes through thirteen years, our analysis is based on the estimation of the effect on the change in employment status, before and after the policy application, for the treatment and control groups.

where the dependent variable is the probability that individual i is unemployed conditional on being employed in the preceding quarter, $\Phi(.)$ is the standard normal cumulative distribution function, D_i is a dummy variable denoting individuals belonging to the treatment group, included on its own and in interaction with the gap between individual i's wage and the new NMW rate, and X_{it} collects all remaining covariates (individual socio-economic characteristics and time effects). An analogous equation can be estimated for the probability of remaining employed conditional on employment in the previous quarter. In line with the standard practice, equation (3), and in particular the coefficient estimate of the first term, is interpreted as capturing the differentiated effect of the minimum-wage increase on the probability of becoming unemployed for the treated individuals relative to those in the control group.

A similar approach can be used to estimate the impact of NMW on the probability of job entry. In this case the equation to estimate is

$$P(e_{t+1} = 1 | e_t = 0) = \Phi(\alpha * D_i + \gamma * X) \quad (4)$$

A particular problem presents itself here in the fact that we do not have any previous wage information for those who enter employment only after the NMW increase. In other words, we do not know whether those entering into employment after the increase would have earned more or less than the minimum wage before the increase. Dickens and Draca (2005) resolve this by defining the treatment group as those whose earnings are less than or equal to the (age-relevant) new NMW rate and the control group as those who earn up to c percent above the NMW:

Treatment group: $w_{t+1} \leq nmw_{t+1}$ (5)

Control group: $nmw_{t+1} < w_{t+1} < nmw_{t+1} * (1 + c)$ (6)

A somewhat uncomfortable implication of this specification is that the treatment group now includes also those who earn less than the NMW (there are specific cases when this is allowed, for example for apprentices or for those who receive employer-provided accommodation or other in-kind payments). An alternative specification would entail constructing the treatment group as including only those who earn the minimum wage after the NMW increase. Using that specification yields very similar results.

Note that our analysis could suffer from a potential endogeneity problem as in a non-experimental sample, such as the Labor Force Survey, workers earning less than the new NMW rate are more likely to lose their jobs even if NMW does not change because they are likely to be less productive than workers earning higher wages.[6] If so, it is the characteristics associated with their lower wages (and not the minimum wage itself) that determine their higher probability of job loss compared to other individuals with above-NMW wages. In other words, if wages are not allocated randomly, the allocation of individuals into treatment and control groups is not random either but depends on their characteristics.

In order to assess to what extent the two groups of individuals are similar, Table 1 presents some basic descriptive statistics for the treatment and control groups, for the analyses of job exit and job entry alike. There are some differences: the individuals in the control groups are slightly more likely to have a university degree or higher education, less likely to have lower qualifications, they are more likely to be white rather than black or Asian, less likely to be a full time student, and they are more likely to live in the rest of South East and South West. However, these differences are generally small and the two groups appear rather similar.

[6] Note however that in expressions (2) and (6), small values of c would imply that the salary of the treated and control groups could be deemed to be very similar.

Table 1. Descriptive Statistics.

	Job entry		Job exit	
	Treatment	Control	Treatment	Control
Higuest qualification				
Degree or equivalent	3.29	4.31	5.04	6.02
Higher education	3.99	4.29	4.7	5.25
GCE A level or equivalent	25.92	25.59	22.58	22.92
GCSE grade A-C or equivalent	33.38	35.35	33.88	35.72
Other qualification	16.24	15.43	18.82	15.81
No qualification	14.55	12.29	12.67	12
Ethnic origin				
White	90.17	91.09	91.91	93.07
Black	6.47	6.23	6.02	5.6
Asian	1.78	1.57	1.98	1.17
Region of usual residence				
Tyne & Wear	3.7	2.36	3.18	2.46
Rest of Northern region	6.47	4.83	5.49	5.2
South Yorkshire	3.62	3.49	3.96	3.43
West Yorkshire	4.11	5.14	5.7	4.76
Rest of Yorks & Humberside	4.11	3.55	4.21	3.99
East Midlands	9.36	10.27	11.06	9.63
East Anglia	3.18	4.2	2.89	2.99
Inner London	1.34	0.92	0.87	1.11
Outer London	2.68	2.78	2.19	2.17
Rest of South East	11.69	14.05	10.44	13.84
South West	7.41	9.35	9	9.7
West Midlands (met county)	5.19	4.06	3.84	4.19
Rest of West Midlands	5.19	5.56	5.16	5.67
Greater Manchester	5.16	4.79	4.21	4.45
Merseyside	3.06	2.2	2.19	2.44
Rest of North West	4.29	4.43	4.46	3.92
Wales	6.82	6.02	6.03	5.67
Strathclyde	4.93	4.56	4.75	3.81
Rest of Scotland	5.45	5.71	5.57	6.69
Northern Ireland	2.24	1.74	4.8	3.9
Whether full time student				
Full time student	19.27	15.03	10.03	8.88
Not full time student	80.35	84.76	89.97	91.12
Sex				
Male	32.48	28.98	27.12	30.14
Female	67.52	71.02	72.88	69.86

Notes: Missing, ‘don’t know’ and ‘other’ responses are not reported.

An alternative approach that overcomes the drawbacks mentioned above is matching: comparing the labor outcomes of the treated individuals with those of similar individuals, with similarity determined based on the set of variables X. Let Y_i be the response, that is $Y_i = 1$ if job exit or job entry is observed and $Y_i = 0$ otherwise; and D_i be an indicator of whether the individual belongs to the treated or control group. In order to compute the causal effect of D_i on the response variable Y, we would need to know the outcome of interest for the same individual if treated, $Y_i(1)$, and if not treated, $Y_i(0)$. However, this is impossible because only one of them can be observed at any given point in time. The counterfactual result, therefore, has to be estimated with a regression model. In this case, we estimate the response Y to a "hypothetical treatment" D_i .

There are two standard approaches to estimate this causal impact. One is to compare the outcome variable of a treated individual with that of one or several other individuals who are as similar as possible to the treated individuals with respect to the values of covariates X_{it}. A second approach matches participants and nonparticipants based on their estimated propensity scores. However, the application of these methodologies is only possible if there is a region of common support between the treatment and control groups.

Regardless of the approach used, the average treatment effect is defined as $\mathrm{ATE} = \mathrm{E}[\mathrm{Y}(1) - \mathrm{Y}(0)]$, where the expected value is computed with respect to the probability distribution of Y for all individuals. We focus on the causal effect for a given set of individuals, for example those who have received the treatment, $\mathrm{E}[\mathrm{Y}(1) - \mathrm{Y}(0)|\mathrm{D} = 1]$, that is, individuals affected by NMW increases. In this case, the expected value is estimated with respect to the conditional distribution of $(\mathrm{Y}|\mathrm{D} = 1)$. Even more generally, if we have a set of covariates X we can estimate the causal effect conditional on them, that is, conditional on $\mathrm{X} = \mathrm{x}$.

However, this is not always possible because matrix X typically has a very high dimensionality and comprises a wide range of covariates, including qualitative and quantitative variables, and some standard approaches such as, for example, the propensity score, cannot be applied if the number of covariates is too high. This forces the analyst to consider a set of variables of lower dimension, putting the strong ignorability assumption in doubt.[7] Besides, the specification of regression models with many variables makes it not practical to consider all possible interactions among the variables. Again, this forces the analyst to consider only interactive effects among first or second order covariates or to use algorithms such as the forward or backward variable selection that may provide locally optimal models. Unfortunately, there is no theoretical justification, only empirical results, to guide us in assessing the scope of a local instead of a global optimum.

Due to these drawbacks, we make use of a particular type of matching estimation based on the BART model for the estimation of causal impact of NMW increases. Being a non-parametric model, this frees us from being restricted by a given model specification. Furthermore, it allows us to estimate with a satisfactory precision the response of the variable of interest to NMW increases, and with that, the counterfactual result even for a high dimensional X. An additional important advantage of this approach is that it allows for identification of the most significant interactive effects between the treatment variable and any of the covariates without being constrained to include these interactions in any parametric form.

In order to assume that the outcome is independent of the treatment, it is necessary to account for all possible conditioning factors by including a broad range of covariates, X. More specifically, the strong ignorability hypothesis with respect to the allocation of treatment states that Y is conditionally independent of D given X and that the probability of treatment allocation is always positive regardless of the specific value of X. However, like in other matching methods, this does not preclude the possibility of selection on unobservables. Under this hypothesis, the estimation of the marginal effects associated to the treatment variable can be considered in general as a consistent and unbiased estimation of the causal effect of NMW on the probability of job exit and job entry: including a relevant set of covariates in equations (3) and (4) is a sufficient condition to ensure an unbiased estimation. However, as argued by Morgan and Winship (2007), the regression approach can be subject to two important drawbacks. The first relates to the fact that

[7] See Caliendo and Kopeining (2008) and references therein for a discussion on this issue.

the causal effect of NMW is not constant across individuals. In this case, the estimated causal effect represent a conditional variance weighted estimate of causal effects of individuals and the causal estimation is only unbiased and consistent for this particularly weighted average that is not usually the parameter of interest. The second problem relates to the fact that the strong ignorability condition does not necessarily imply that treatment is uncorrelated with the error term net of adjustment for X as this error term depends on the specification of covariates, X. Therefore, in order to interpret the estimation of a regression strategy as a reliable causal effect, we require a fully flexible parameterization of X.

4 BART Model

In the following explanation of the model, we mainly follow the notation of Hill (2011) and references therein (see Chipman et al., 2010, for details of the statistical model, and Leonti et al., 2011, for an application of this model to the estimation of a causal effect of the use of medical plants). Let $\mathcal{D}ata$ be the available data, that is the set Y, X, D observed for N individuals and $\pi(\cdot \mid \cdot)$ the probability distribution of the left argument conditional to the right argument. The aim of the analysis is to estimate the *posterior* probability distribution of the causal effect, that is $\pi(ATE|\mathcal{D}ata)$, or the same posterior distribution but conditionally on some covariates, $\pi(ATE|\mathcal{D}ata, X = x)$. In order to do this we use a non parametric regression model. The novelty in these types of causal inference analyses is the use of a Bayesian regression model known as BART. As in all Bayesian models, we need a likelihood function defined for a set of parameters, $\theta \in \Theta \notin \mathbb{R}$, and a prior distribution $\pi(\theta)$, $\theta \in \Theta$. The likelihood function, $L(Y|X, D, \theta)$, is obtained from the following additive regression model, where the mean of Y, $E(Y) = P_i(Y = 1) = P$, is determined from the sum of estimated models for the response variable:

$$P = \Phi\left[\sum_{j=1}^{m} g\left(X, D; T_j, M_j\right)\right] \quad (7)$$

where $g\left(X, D; T_j, M_j\right)$ is a classification tree with the variables and split points represented by T_j and the terminal nodes denoted by M_j and computed with respect to the values x, D that belong to the individual whose response is Y. Essentially, g is a function that gives to each individual i

their expected value in the j^{th} tree, $\mu_{ij} \in M_j$. This part of the model, $\sum_{j=1}^{m} g\big(X, D; T_j, M_j\big)$, operates in the same way as the usual linear predictor in an ordinary regression model, in fact if we substitute the sum of trees with the usual linear predictor, we were dealing with the ordinary linear regression model with the least square estimator. Of course the problem at hand needs a very flexible model, which would be very difficult to be obtained with a linear predictor. Essentially, viewing by T_j and the terminal nodes denoted by M_j as model parameters, we allow the data to define the terms that enter into this kind of linear predictor instead being fixed beforehand by the analyst. The final score estimated for the i^{th} individual would correspond to the average of the *m* scores over all trees in which each tree has been grown in order to capture a specific aspect of the relation between the response is Y and the rest of predictors. It is well known that, in order to minimize the forecast error, classification trees tend to grow disproportionally until generating *overfitting* in the response and that in general an estimator obtained from many simple trees is more efficient than another one obtained from a single complex tree. Examples of these types of models are Boosting (Shapire and Singer, 1999) and Random Forest (Breiman, 2001).

In order to achieve this necessary tree simplification, we use a regularization prior on the size of the tree $\pi(T, M)$ as specified in Chipman et al (2010). This regularization prior precludes the tree from growing too much and makes sure that each of the μ_{ij} contributes in a marginal way to the estimation of the response function. As Chipman et al (2010) show, the hyper parameters of all prior distributions are specified in relation to the observed sample. It produces priors that are dependent on the sample. This procedure, which is not very orthodox from a Bayesian point of view, is part of the approaches known as empirical Bayes that are very popular and have been enhanced from a theoretical point of view by Petrone et al. (2013). As explained by Hill (2011), the results of this type of analysis are robust with respect to prior modifications.

Using the priors specified above it is possible to simulate samples of the posterior distribution with a non-excessive computational effort using Markov Chain Monte Carlo (MCMC), more specifically using Metropolis Hastings. In particular, the proposal distribution used in the MCMC to update the values of T_j and M_j consists of adding/dropping a terminal node and changing a split variable or a split point with the probabilities specified in Chipman et al. (2010). Such probabilities, which finally define the proposal in the MCMC scheme, are set according to the observations in order to guarantee an optimal mixing of the chain and so increase the precision in the posterior estimation. Once the posterior distribution of $\theta = (T_1, \dots, T_m, M_1, \dots, M_m)$ has been obtained, the predictive distribution for the probability of job exit is:

$$m(Y_i|x_i, z_i) = \int_{\theta \in \Theta} L(Y_i; \theta) d\pi(\theta|\pi(\theta)) \quad (8)$$

where L is the likelihood function for $\theta \in \Theta$ and $\pi(\theta)$ is the posterior. Integral (8) is practically estimated by generating values of $P_i = P_i(Y_i = 1)$, using the normal distribution with the mean and variance for each value θ in the chain MCMC and the regression tress computed using the values of regressors for individual i , that is x_i and z_i. In particular we use m=500 trees and 5000 MCMC steps after an initial burn-in of 1000 steps.

In this way, the distribution for each individual and the corresponding counterfactual response can be estimated simply by estimating the response in $\mathrm{D_i} = 1$ if the worker is affected by NMW and in $\mathrm{D_i} = 0$ otherwise. Once these predictive posterior distributions have been obtained, the difference between the factual and counterfactual responses are considered to obtain the distribution of the individual causal effect. Finally, $\pi(\mathrm{ATE}\backslash\Omega)$ is estimated from the set of the differences for all the individuals. Finally, the estimation of the conditional causal effect is obtained simply by considering the difference for the individuals that fulfill the condition $\mathrm{X} = \mathrm{x}$.

5 Results

As a first step, and to establish a benchmark to compare our results against, we report the results of a probit model as specified in Equations (3) and (4), where job entry and job exit are functions of the dummy variable for the treatment along with a set of covariates (Table 2).[8] In this regression, the parameter c defined in the previous section is set to be 0.1 to ensure that treatment

[8] Besides standard socio-economic characteristics, we also include an indicator to account for the fact that the age limit for the adult rate was lowered from 22 to 21 from October 2010.

and control individuals are comparable in terms of wages but the results are qualitatively similar when we consider $c = 0.3$, $c = 0.5$ and $c = 1$. The last row of Table 2 indicates that the probabilities of job entry and job exit are both positively correlated with being in the treatment group.

It is interesting to compare these results with those obtained with a standard matching procedure such as the propensity score. The estimation results are qualitatively, and even quantitatively, similar to those obtained from a regression probit model. More specifically, the estimated causal impact for job entry is 0.051 with standard deviation 0.013 while that for job exit is 0.03 with standard deviation 0.011.[9] The fact that the two sets of results are very similar is not surprising as the matching estimation can be interpreted as being similar to a regression that puts more weight on the observations in the treatment and control groups that are very similar to each other.

Table 2. Probability of job entry and job exit as a function of treatment and change.

	Probit regression	
	Job entry	Job exit
LR Chi2	445.13	400.17
Prob > chi2	0.0000	0.0000
Number of obs	7792	6746
Treatment	.0517764**	.024584**
	(.00961)	(.0086)

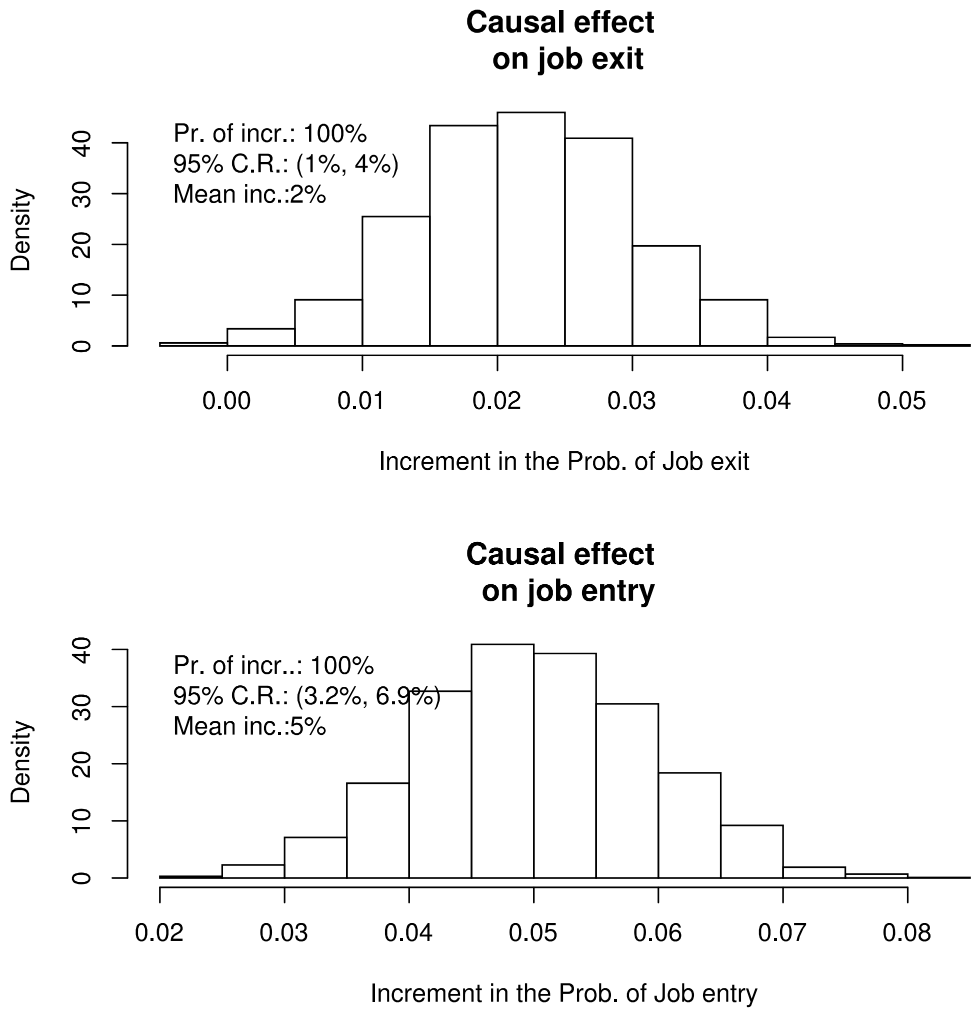
Notes: Marginal effects evaluated at mean values. Significance: ** 1%, * 5%.

The Bayesian approach considered here is instead based on the estimation of the expected value of the treatment and control groups using the same explanatory variables in both cases. Figure 1 reports the estimated distributions of the total causal impact of increases in the minimum wage rate on job exit and job entry using the BART model with all workers aged 18-40. The results indicate that the treatment has positive effects on both job entry and job exit, in a manner similar to the probit results reported above. More specifically, the NMW exerts a positive impact on job entry, and the mean value of this causal impact is 5% with a 95% confidence interval of [3.2%, 6.9%]. For job exit, the effect is positive with the mean value equal to 2% and with a 95% confidence interval of [1%, 4%]. Three cautionary notes are required here. First, the

[9] These results are available upon request.

aforementioned effects are only measured for those workers actually affected by the NMW increase. The minimum-wage increases only affect low-paid workers and need not apply throughout the distribution of wages. Second, although the estimated effect on job entry is larger than that for job exit, the overall effect of NMW is ambiguous as there are many more employed workers (who are candidates for job exit) than unemployed individuals (candidates for job entry). Third, it is possible that NMW increases have spillover effects whereby wages just above the new minimum wage also increase. The available literature suggests that such spillovers are small or none in the UK (Dickens and Manning, 2004; Steward, 2011). If present, such spillovers would bias the estimated effects downwards.

Figure 1 Posterior distribution of the causal effect of job entry and job exit



As discussed above, one of the most important advantages of the BART approach is that it allows for the simultaneous estimation of any kind of interaction between the treatment variable and any of the covariates. This is possible either at model estimation or at description level of the obtained results. Here we consider the result at the description level by inspecting the interaction between covariates and the estimated causal effect. In particular, the interaction with categorical variables is evaluated trough boxplots, which include 95% percentile bootstrap confidence intervals for the median, while the interactions with continuous covariates by local polynomial regression smoother (loess) along with their 95% confidence intervals (Cleveland et al., 1992, Chp. 8) .

In Figure 2, we present the interaction between the NMW increase and the size of the increase. While the effect of size is significant, no systematic pattern can be discerned: the estimated effects oscillate around the mean values reported in Figure 1, neither increasing nor decreasing as the size of the NMW change goes up. Next, we interact gender with the effect of NMW increases (Figure 3). Again, the previous finding of a greater effect of NMW increases on job exit than on job entry is reproduced. Although for job entry it is clear that the median values are significantly different, the distributions of the two effects are very similar which suggests that gender plays little role. In Figure 4, in turn, we consider the interaction with age (expressed in months rather than years). Here, the pattern is different for job exit and entry. While the causal impact of NMW is decreasing with age in both cases, that decline is much steeper for job entry. This is not surprising, given that young workers are more vulnerable to NMW increases. Besides, the interactive effect is clearly stronger for job entry. In Figure 5, we consider the interaction with the highest attained qualification. Again, although it is possible to observe significantly different mean values associated to the different qualifications, the whole distribution of the estimated causal effect indicates that this variable is not a relevant factor to explain differences in the causal impact of NMW either for job entry or job exit. Finally, Figure 6 presents the interaction with the regional business cycle – measured using the unemployment rate. Interestingly, this interaction effect is very different for the two labor-market flows: the minimum-wage effect on job exit is relatively low and depends little on the regional unemployment rate, whereas that for job entry is higher and positively related to regional unemployment. This implies that the effect of the minimum wage on job entry differs considerably between recessions and booms, whereas the business cycle has little bearing on how the minimum wage shapes job exits.

So far we have been considering the effects of NMW changes for two similar groups, those affected by the change and those who are unaffected but are otherwise similar to the affected individuals both in terms of their wage and in terms of the other covariates used in the analysis. However, to test for the robustness of our results even further, we carry out a falsification experiment whereby we define the treatment and control groups as if the NMW were equal to the actual NMW plus 2£. Our hypothesis is that neither job entry nor job exit should be affected by a wrongly-defined increase in the NMW. The results of this experiment, shown in Figure 7, indicate that the causal impact of the (false) NMW increase is significant at the 5% level for job entry but not for job exit. We find similar conclusions for other artificial NMW rates (results are available from the authors upon request). Importantly, the insignificant falsification test results for job exit give strong support to the finding that the employment of workers earning the minimum wage is adversely affected by NMW increases.

The fact that the falsification test is significant for job entry decisions could be due to potential unobservable variables not included in the model. Another possibility is that it is driven by spillover effects of the actual NMW increase: the NMW change can lead to ripple effects for wage rates above the minimum wage.[10] To account for this possibility, we consider an alternative definition of the control group:

$$\mathrm{nmw}_{t+1} * (1 + \mathrm{a}) \leq \mathrm{w}_{\mathrm{it}} < \mathrm{nmw}_{t+1} * (1 + \mathrm{c}) * (1 + \mathrm{a} + c) \qquad (9)$$

[10] As long as the direct effect of the NMW increase is larger than the indirect (spill-over) effect, we can obtain a significant result. Otherwise, it would be impossible to define control and treatment groups. As discussed above, the available evidence so far suggests that such spillovers are limited or zero (Dickens and Manning, 2004; Steward, 2011).

where a = 0.1. The treatment group is defined as before (see equation 2). The new definition ensures that the control and treatment groups are close enough but are not immediately adjacent to each other in the distribution of wages. The results with the alternatively defined control group, presented in Figure 8, confirm the previous results: the mean effect is the same for job exit and is only slightly higher for job entry with the new control group. This suggests that the spillovers are very limited, if any.

Figure 2 NMW Increases, Job Exit/Entry and Size of the Increase

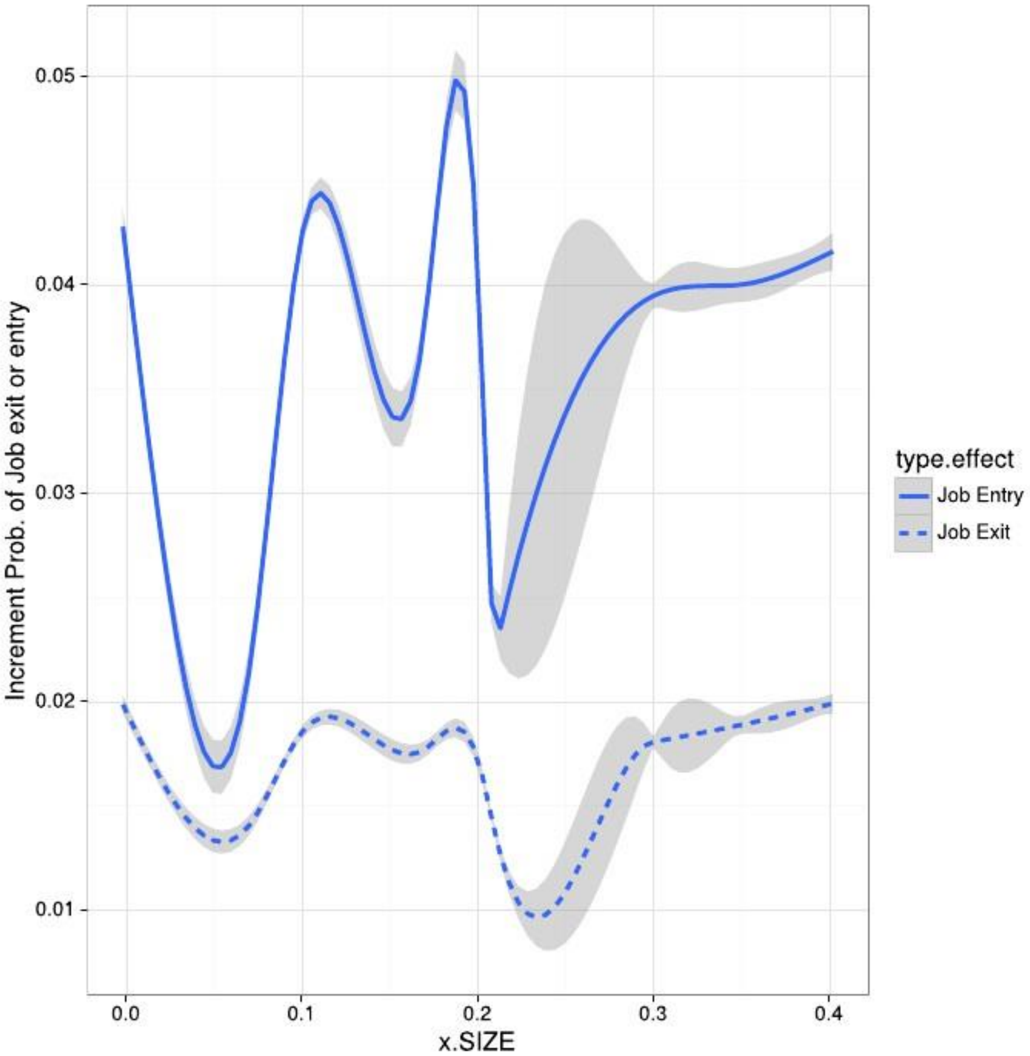


Figure 3 NMW Increases, Job Exit/Entry and Gender

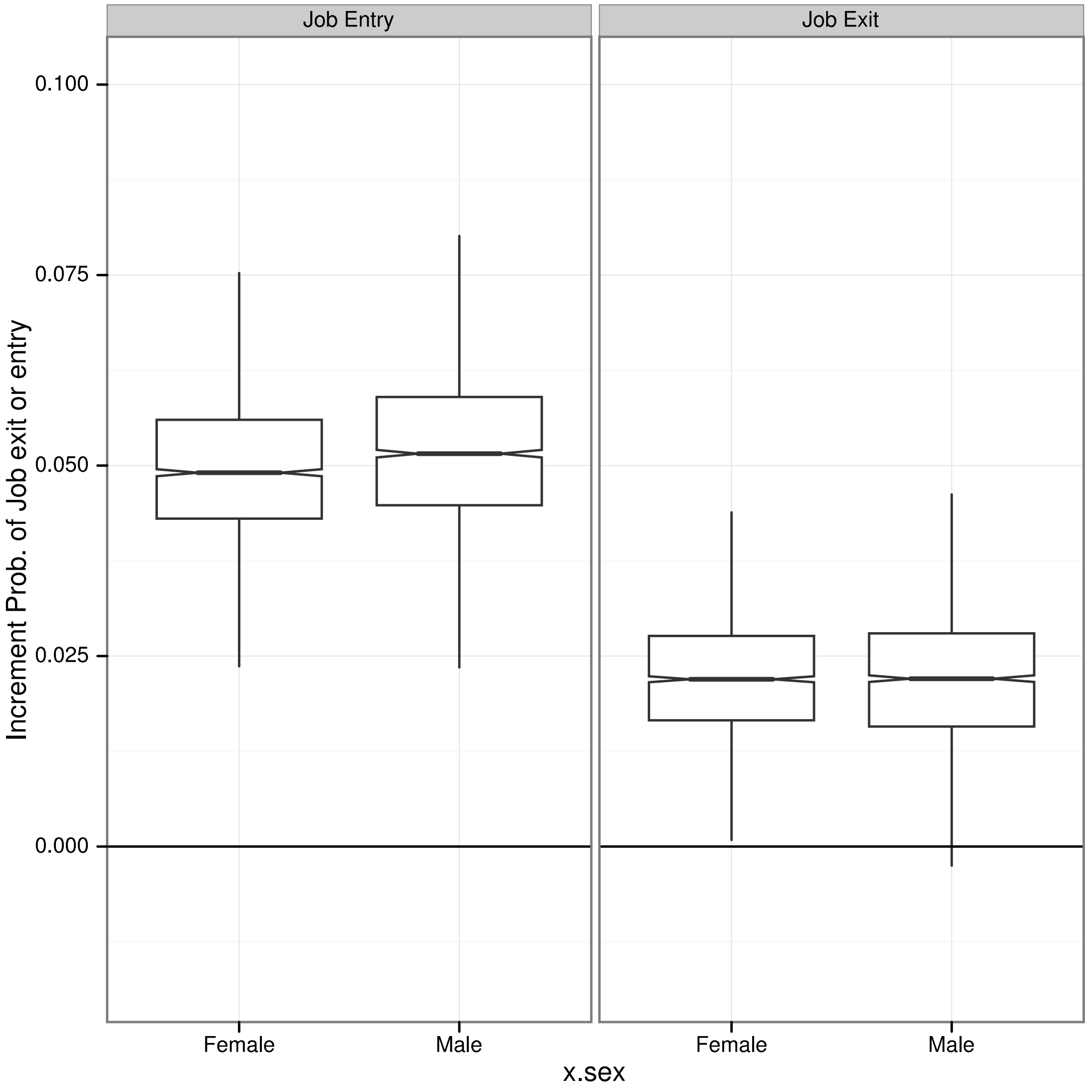
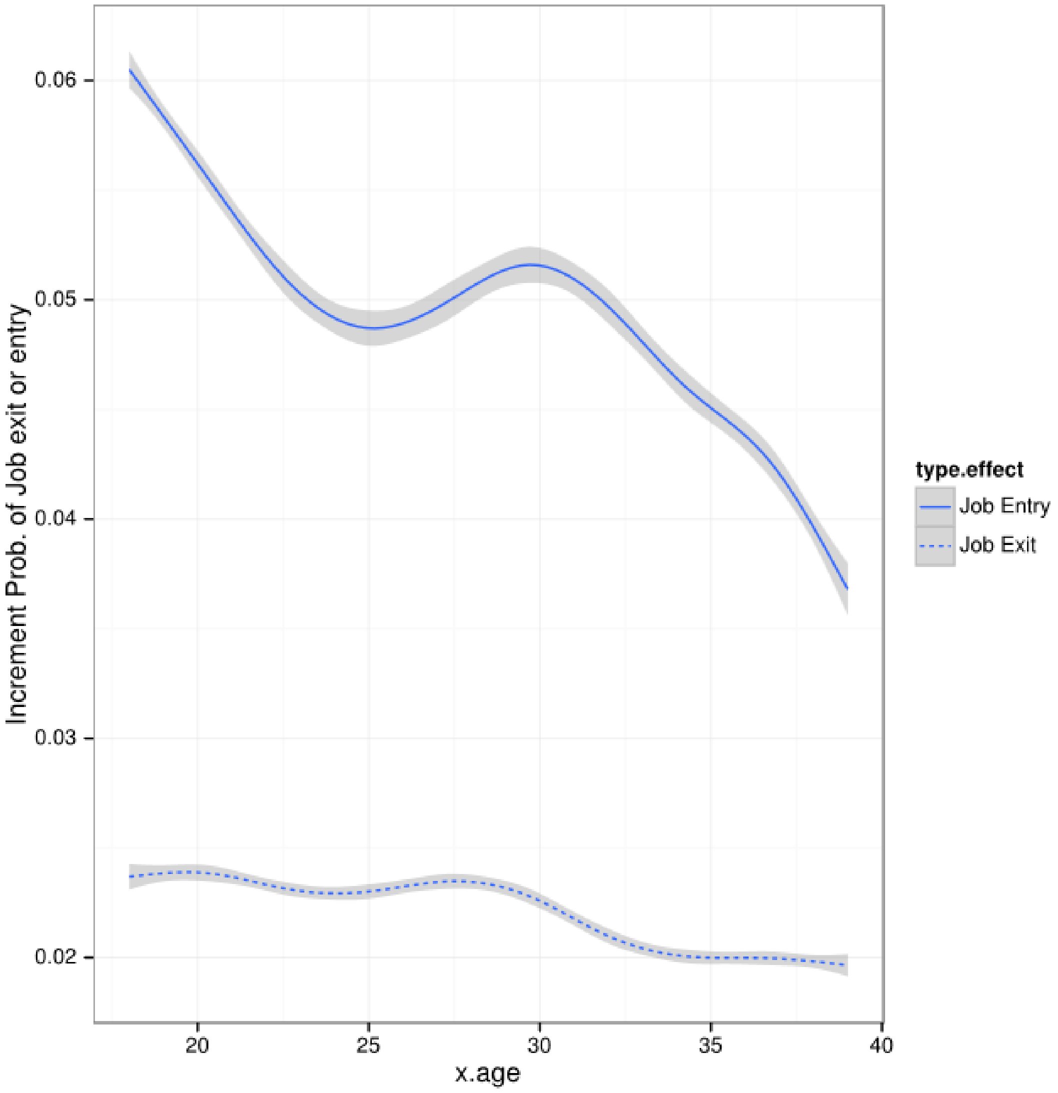
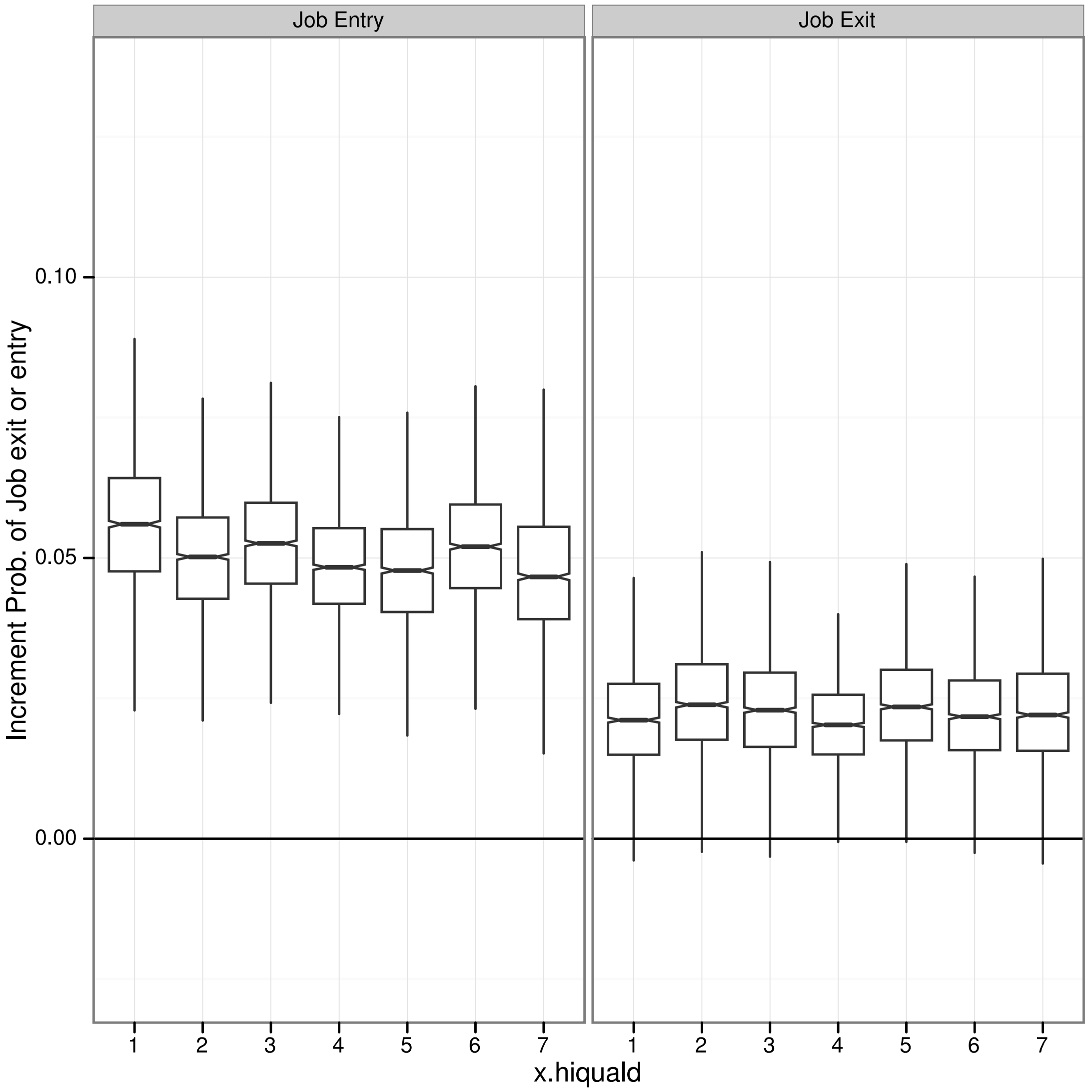


Figure 4 NMW Increases, Job Exit/Entry and Age



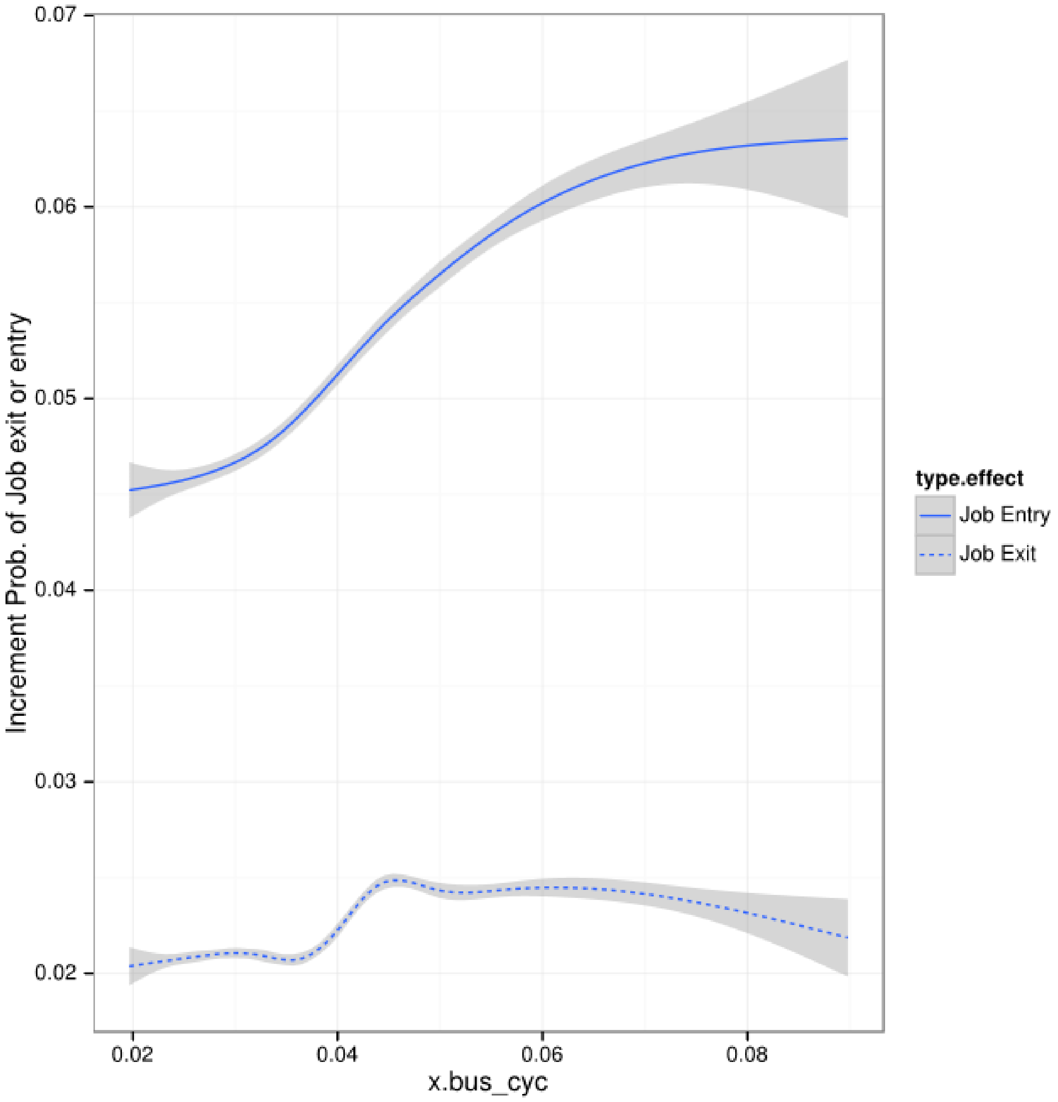
Note: Shadow area indicates the 95% confidence interval of the local polynomial regression estimator (loess).

Figure 5 NMW Increases, Job Exit/Entry and Qualifications



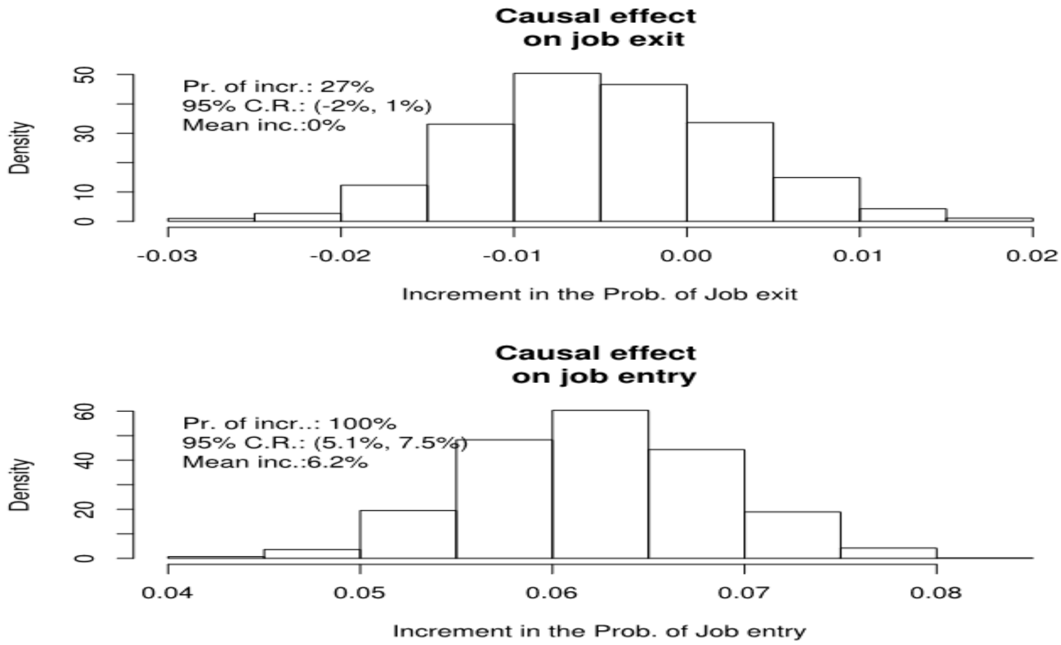
Notes: 1 Degree or equivalent, 2 Higher education, 3 GCE A Level or equivalent, 4 GCSE grades A-C or equivalent, 5 Other qualifications, 6 No qualification, 7 Don’t know

Figure 6 NMW Increases, Job Exit/Entry and Business Cycle



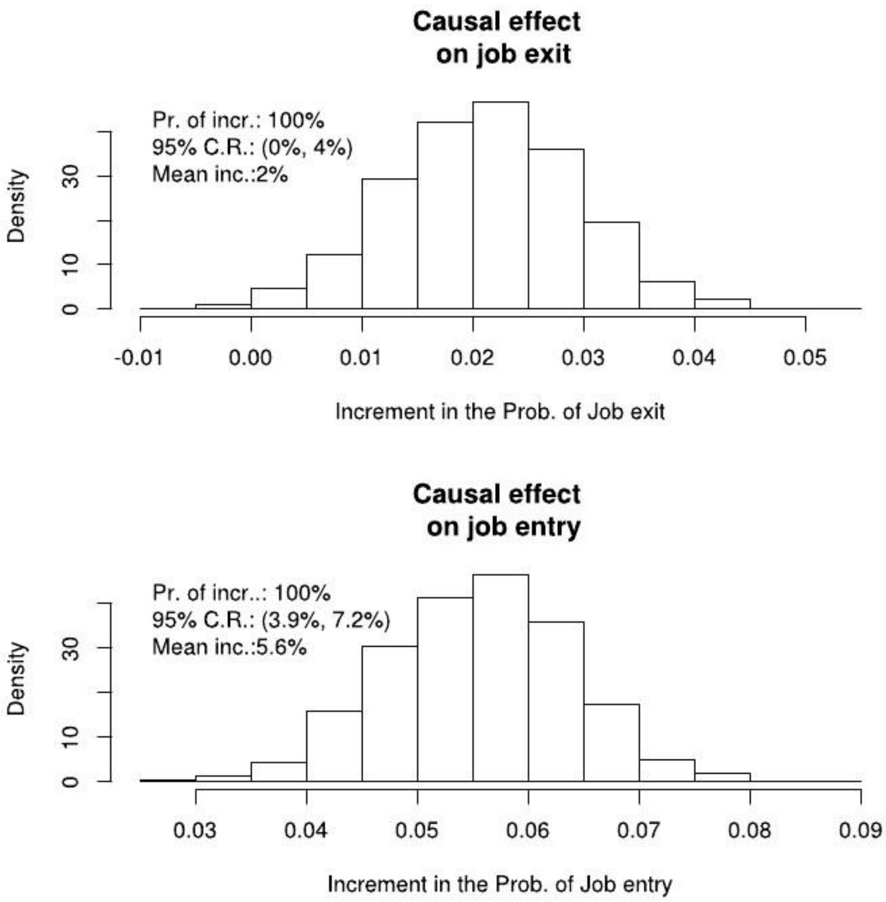
Notes: The horizontal axis measures the regional unemployment rate. Shadow area indicates the 95% confidence interval of the local polynomial regression estimator (loess).

Figure 7 Falsification Test: NMW + £2



Notes: The falsification test simulates the NMW being £2 higher than the actual value.

Figure 8 Alternative Control Group



6 Concluding remarks

We estimate the causal impact of the NMW on the probability of job entry and job exit in the UK, applying a novel methodology to this context, the Bayesian Additive Regression Trees (BART). An important advantage of this procedure is that it allows the identification of the most important interactions between the treatment variable and other covariates in the model. We find that the NMW exerts a significantly positive effect both on job entry and job exit, with the impact on job entry being relatively stronger (given that there are fewer unemployed than employed workers, the absolute size of the flows cannot be readily compared). The causal effect of NMW is found to be higher for young workers and in periods of high unemployment; both of these interactions are more prominent for job entry than for job exit. However, no significant interactions were found with gender and worker qualification. Overall, the effect of the NMW on low-paid workers is stronger for job entry than for job exit.

Most importantly, our paper open new lines of research that can be explored in subsequent work. For example, this fully flexible approach could be adapted to deal with some recent issues in the literature about the importance of the econometric specification on estimating the effect of minimum wage using panel data models in US states. Also, it could be used to estimate the possible interactions between the federal minimum wage and the state minimum wages, as done by Baskaya and Rubinstein (2012), without the necessity of estimating two different models.

References

Abadie, A., and G.W. Imbens (2006). “Large Sample Properties of Matching Estimators for Average Treatment Effects,” *Econometrica* 74(1), 235-267.

Allegretto, S., Dube, A., Reich, M., (2011), “Do minimum wages really reduce teen employment? accounting for heterogeneity and selectivity in state panel data.” *Industrial Relations* 50, 205-240.

Allegretto, S., Dube, A., Reich, M. and Zipperer, B., (2013), “Credible research design fo minimum wage studies.” Working paper, Institute for Research on Labor and Employment, UC Berkeley.

Baskaya, Y., Rubinstein, Y., (2012), “Using federal minimum wages to identify the impact of minimum wages on employment and earnings across the U.S. States.” Unpublished paper, Central Bank of Turkey.

Breiman, L. (2001). “Random forests”. *Machine Learning* 45, 5-32.

Card, D. and Krueger, A.B. (1995), “Time-series minimum wage studies: a meta-analysis.” *American Economic Review* 85, 238-243.

Caliendo, M. and Kopeining, S. (2008), “Some practical guidance for the implementation of propensity score matching.” *Journal of Economic Surveys* 22, 31-72.

Chipman H.A. , George E.I., and Mculloch R. (2010). “BART: Bayesian Additive Regression Trees”, *Annals of Applied Statistics* 4(1), 266–298.

Cleveland W. S., Grosse E. and Shyu W. M. (1992) Local regression models. Chapter 8 of Statistical Models in S eds J.M. Chambers and T.J. Hastie, Wadsworth & Brooks/Cole.

De Linde Leonard, M., Stanley., T. D. and Doucouliagos, H. (2014). “Does the UK Minimum Wage Reduce Employment? A Meta-Regression Analysis.” *British Journal of Industrial Relations* 52(3), 499–520.

Dickens, R. and M. Draca (2005), “The Employment Effects of the October 2003 Increase in the National Minimum Wage,” CEP Discussion Paper No 693, Centre for Economic Performance, London School of Economics.

Dickens, R., and A. Manning (2004). “Spikes and Spillovers: The Impact of the National Minimum Wage on the Wage Distribution in a Low-Wage Sector.” *Economic Journal* 114, C95–C101.

Doucouliagos, H. and Stanley, T.D. (2009), “Publication selection bias in minimum-wage research? a meta-regression analysis.” *British Journal of Industrial Relations* 47, 406-428.

Dube, A., Lester, T.W., and Reich, M. (2010), “Minimum wage across state borders: estimates using contiguous counties.” *Review of Economics and Statistics* 92, 945-64.

Hill, L.H. (2011). “Bayesian Nonparametric Modeling for Causal Inference”, *Journal of Computational and Graphical Statistics* 20(1), 217-240.

Lee, K.C,, Pesaran, M.H. and Pierse, R.G. (1990), "Testing for Aggregation Bias in Linear Models." *The Economic Journal* 100(400), 137-150.

Leonti, M., S. Cabras, C. Weckerle, M. Solinas, and C. L. (2010). “The causal dependence of present plant knowledge on herbals—contemporary medicinal plant use in campania (italy) compared to matthioli (1568)”, *Journal of Ethnopharmacology* 130 (2), 379–391.

Hirsch, B.T., Kaufmann, B.E., and Zelenska, T. (2015). “Minimum Wage Channels of Adjustment.” *Industrial Relations* 54(2), 199-239.

Morgan, S.L., and C. Winship (2007), *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press.

Neumark, D. and Wascher, W. (2007), “Minimum wages on employment.” *Foundations and Trends in Macroeconomics* 3, 1-182.

Neumark, D., Salas, I.M.I, and Wascher, W. (2014), "Revisiting the minimum wage employment debate: throwing out the baby with the bathwater?" *Industrial and Labor Relations Review* 67, 608-648.

Petrone S, Rousseau J, and Scriocciolo S. (2014). "Bayes and empirical Bayes: Do they merge?" *Biometrika* (to appear).

Sabia, J.J. (2014), "Minimum Wages and the Business Cycle: Does a Wage Hike Hurt More in a Weak Economy?" Employment Policies Institute: Washington, DC.

Sparapani, R.A., Logan, B.R., McCulloch, R.E. and Laud, P.W. (2016). "Nonparametric survival analysis using Bayesian Additive Regression Trees (BART)." *Statistics in Medicine* 35(16), 2741-2753.

Schapire, R.E. and Singer, Y. (1999). "Improved boosting algorithms using confidence-rated predictions", *Machine Learning* 37, 297-336.

Stewart, M.B. (2004). "The employment effects of the national minimum wage", *Economic Journal* 114, 110-116.

Stewart, M.B. (2011). "Wage Inequality, Minimum Wage Effects and Spillovers." Warwick Economic Research Paper 965.

Tan, Y.V., Flannagan, C.A.C., and Elliott, M.R. (2016). "Predicting human-driving behavior to help driverless vehicles drive: random intercept Bayesian Additive Regression Trees." Available online: https://arxiv.org/pdf/1609.07464v1.pdf