Brunel University London

Department of
Economics and Finance

Economics and Finance Working Paper Series

Working Paper No. 14-22

Stefano Cabras, Jan Fidrmuc and J.D. Tena

Minimum Wage and Employment: Escaping from the Parametric Straitjacket

December 2014

http://www.brunel.ac.uk/economics

Minimum Wage and Employment: Escaping from the Parametric Straitjacket

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

December 2014

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). This procedure is simple, flexible and it does not require ad-hoc assumptions about model fitting, number of covariates and 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 lot of more employed workers. This can explain the insignificant effect found in the previous studies based on aggregate macroeconomic data. Furthermore, the causal effect of NMW is higher for young workers and in periods of high unemployment and they have a stronger impact on job entry decisions. However, no significant interactions were found with gender and worker qualifications.

Keywords: BART, causal inference, difference-in-difference, matching regression.

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

[*] 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.

‡ Departamento de Estadística, Universidad Carlos III and Dipartamento di Economia e Impresa, Università di Sassari. Contact information: C/Madrid 126. 28903 Getafe (Madrid), Spain, jtena@est-econ.uc3m.es. Phone: +34 916241259 .

1 Introduction

The most characteristic feature regarding the literature on the causal impact of the minimum wage on employment is the complete 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, a meta-analysis on this issue conducted by Card and Krueger (1995) and the subsequent correction by Doucouliagos and Stanley (2008) both find that although there is a wide range of results in the previous research, the overall mean is consistent with a non significant impact of the minimum wage on employment output once the publication selection bias is accounted for.

The 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 highly 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. Therefore, 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) 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 critical determinant of whether a significant or insignificant impact of minimum wage on state employment is obtained.

Aggregation bias can be a potential second problem of the literature discussed above as an aggregate data might mask the real effect of minimum wage at the individual level. This paper uses 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). 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 with a similar individual whose salary is not affected but it is sufficiently close to those in the treatment group. However, this procedure has important advantages over other more traditional parametric specifications. Among them, it does not require any type of hypothesis over the covariates to be included in the model, the ability to consider a large number of regressors and to estimate any type of interactive effects between the treatment variable and any other variable in the analysis, or the way this methodology copes with missing values without the necessity of dropping them from the sample. Thus, under the BART model, for each treated individual the definition of the closest untreated individual and the interactions between the different clusters of individuals and time or and any other relevant covariate is not constrained to follow any ad-hoc parametric function.

This paper is closely related to at least three previous works that estimate the impact of the NMW on employment at the individual level using micro-data. Stewart (2004) and Dickens and Draca (2005) study the impact of the NMW by applying the difference-in-difference regression techniques to the UK Labor Force Survey data, finding that the NMW does not have a significant adverse effect on employment. In both cases, the treated group are workers whose salary must be modified due to the NMW change while the control group are workers whose salary is not affected by the NMW change but it is very close to those affected by it. In particular, Stewart (2004) analyzes how the introduction of the UK NMW in 1999 and their subsequent changes in 2000 and 2001 has 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. For the US, Sabia et al. (2013) also use a difference-in-difference approach to estimate the effect of an increase in the minimum wage in New York on low skilled individuals' employment. They find that the minimum wage has a negative impact on employment that is robust to using different control groups: neighbor states, high skilled workers or a synthetic control group. Unlike these papers, we do not study the impact of NMW for a specific number of years but the impact of all the NMW changes since its introduction in 1999. Moreover, we follow a matching estimation that can be interpreted as being similar to a regression that puts more weight on observations that are very similar to the treatment and control groups. As we will discuss later, this has important advantages for the reliability of the estimation results. Finally, the approach used in this paper allows us to identify the most relevant interactions of the NMW effect with other variables such as gender, age, qualifications and business cycle without the necessity of proposing a parametric specification.

To preview the results, we find that the NMW exerts a positively 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 a lot of more employed workers than the number of the unemployed. This could explain the insignificant effect found in the previous work based on aggregate macroeconomic estimations. We find that the causal impact interacts with workers' age and business cycle such that the effect is stronger for younger workers and in high unemployment periods. Other variables such as gender or qualifications do play an important role in these interactions.

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 Labour Force Survey (LFS). The LFS is a quarterly nationally-representative survey of households across the UK. Each quarterly file contains information on approximately 60 thousand households and over 100 thousand individuals aged 16 and above. 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 labour market outcomes. Since the NMW was introduced in April 1999, we use all quarterly datasets available since 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[1]), and the adult rate.[2] Historically, the youth rate was set some 35% higher than the 16-17 rate while the adult rate exceeds 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.[3] 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. Furthermore, we do not analyze the effects of differences between the different age-specific NMW rates but instead consider the regular annual increases that apply to all rates.

[1] The upper limit for the youth rate has been lowered to 20 from October 2010. Where relevant, our analysis takes this change into account.

[2] 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.

[3] 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 national minimum wage increases on employment by going beyond the difference-in-difference and the matching estimation methodologies traditionally used for this purpose. We start by explaining the difference-in-difference and matching models and then outline why and how our approach differs from them.

The difference-in-difference methodology involves comparing the changes in the outcomes (such as employment) after a NMW change for the treatment and control groups. 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:

$$\mathrm{nmw_t} < \mathrm{w_{it}} < \mathrm{nmw_{t+1}} \quad (1)$$

where $\mathrm{nmw_t}$ is the (age-dependent) NMW rate in effect at time t while $\mathrm{w_{it}}$ is the worker'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 are more likely to possess similar characteristics as those earning the minimum wage. If we set the upper bound as *c* above the new rate, the control group comprises workers meeting the following condition:

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

We can then estimate the following equation

$$P(e_{t+1} = 0 | e_t = 1) = \alpha * D_i + \gamma * X_{it} \qquad (2)$$

where the dependent variable is the probability that individual i is unemployed conditional on being employed in the preceding quarter, 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 (1), and in particular the coefficient estimates of the first two terms, are 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

$$P(e_{t+1} = 1 | e_t = 0) = \alpha * D_i + \gamma * X \qquad (3)$$

A particular problem presents itself here in the fact that we do not have any previous wage information for those who only enter into employment 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) define 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: $w_{t+1} \leq nmw_{t+1}$ (4)

Control group: $nmw_{t+1} < w_{t+1} < nmw_{t+1} * (1 + c)$ (5)

A somewhat uncomfortable implication of this specification is that the treatment group now includes also those who earn less than the NMW (recall that there are specific cases when this is allowed, for example for apprentices). An alternative specification would entail constructing the treatment group as including only those who earn the minimum wage after the NMW increase.

Note that in a non-experimental sample, such as the Labor Force Survey, the outcomes of interest are not typically independent of the treatment. This is known in the literature as the endogeneity problem. In our particular case, workers earning less than the new NMW rate are expected to be more likely to lose their jobs even if NMW does not change. Therefore, 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 assume that the outcome is independent of the treatment, it is necessary to account for all possible conditioning factors by the 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. 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. Although including a relevant set of covariates in equations (2) and (3) is a sufficient condition to ensure an unbiased estimation, however, as indicated by Morgan and Winship (2007), the regression approach can be subject to two important drawbacks. The first one relates to the fact that the causal effect of NMW is not constant over 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 with the regression strategy to estimate the causal effects relate 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 real causal effect, we require a fully flexible parameterization of X.

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 probability of job exit 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 should know, in principle, 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 mainly 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 nonparticipants as similar as possible to the participant considering 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, once the potential causal effects have been estimated the average total 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 the individuals. The causal effect for each individual is of no interest but what is of interest is 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, quantitative and sortable variables and the treatment and non-treatment cannot be observed for the same value of $\mathrm{X} = \mathrm{x}$.

Moreover, 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.[4] Finally, 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 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.

4 BART Model

In the 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 Ω 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 $\pi(ATE|\Omega)$, or the distribution conditional on some covariates, $\pi(ATE|\Omega, X = x)$. In order to do this we use a non parametric regression model. The novelty in these types of causal inference

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

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 is determined from the sum of estimated models for the response variable:

$$Y = \Phi\left[\sum_{j=1}^{m} g\left(X, D; T_j, M_j\right)\right] \quad (6)$$

where $\Phi[.]$ is the standard normal cdf and $g\left(X, D; T_j, M_j\right)$ 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 P. Essentially, g is a function that gives to each individual i its expected value in the jth tree, $\mu_{ij} \in M_j$. The final score estimated for the ith individual would correspond to the average of the m scores. 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, it is necessary to use a regularization prior on the size of the tree $\pi(T, M)$ 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. The posterior distribution of θ is estimated in a computationally feasible way by considering a conjugate prior on σ^2, that is, an inverse-gamma that induces a conditional distribution of σ^2, $\pi(\sigma^2 | T_1, \dots, T_m, M_1, \dots, M_m)$ has a closed form expression that is again an inverse-gamma. 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 a recent paper 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 within Gibbs. This means that the simulation algorithm alternates Gibbs steps (as the one that is necessary to simulate σ^2) and Metropolis Hastings steps when the conditional distributions for the remaining parameters are not available in a closed form expression. In particular, the distribution used 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). Once the posterior distribution of $\theta = (T_1, \dots, T_m, M_1, \dots, M_m, \sigma^2)$ has been obtained, the predictive distribution for the probability of job exit is:

$$m(Y_i|x_i, z_i) = \int_{\theta\in\Theta} L(Y_i;\theta) d\pi(\theta|\Omega) \tag{7}$$

that is practically estimated generating values of P_i, using the normal distribution with the mean and variance for each value θ in the chain MCMC and the regression tress computed in 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. Then, the estimation of the conditional causal effect is required, this is obtained simply by considering the difference for the individuals that fulfill the condition $\mathrm{X} = \mathrm{x}$.

5 Results

As an initial analysis, we show in Table 1 the results of a probit model for job entry and job exit as specified in Equations (1) and (2) as a function of the dummy variable for the treatment along with and a set of covariates.[5] In this regression c is set to be 0.1 but the results are qualitatively similar when we consider $c = 0.3$, $c = 0.5$ and $c = 1$. The last row of Table 1 indicates that the probabilities of job entry and job exit are both positively correlated with being in the treatment group.

Table 1. 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.

It is interesting to compare these results with those obtained with a standard matching estimation 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. This is not surprising as a 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.

To address this possibility, Table 2 reports the average values for the main variables for the treatment and control groups. Clearly, the distributions of explanatory variables are not the same. Although these differences are small and not significant at the conventional level for each individual variable, when they are jointly considered it cannot be appraised if they may have an important impact on the estimated causal effect. Moreover, the more variables are included in

[5] 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.

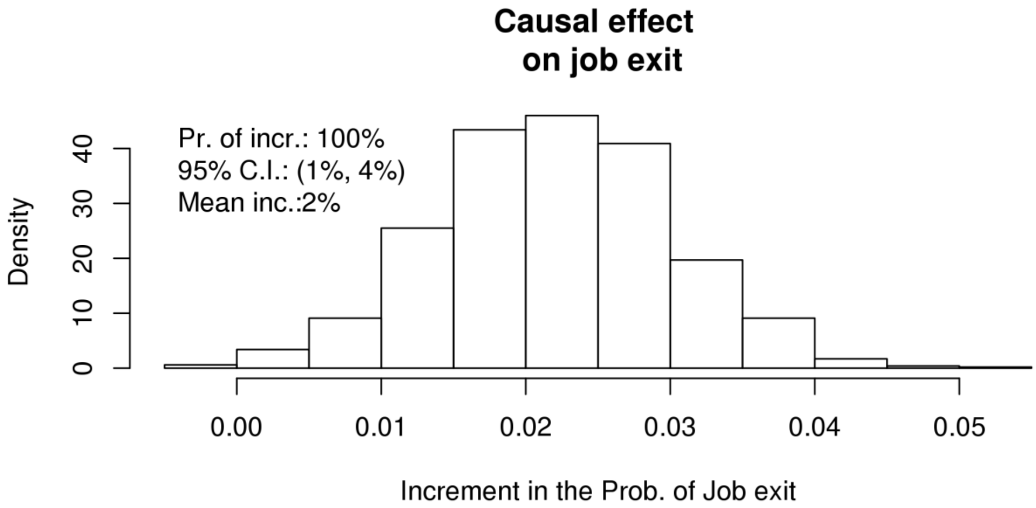
the analysis the more difficult this comparison becomes as each new variable introduces new sources of divergences between the treatment and control groups.

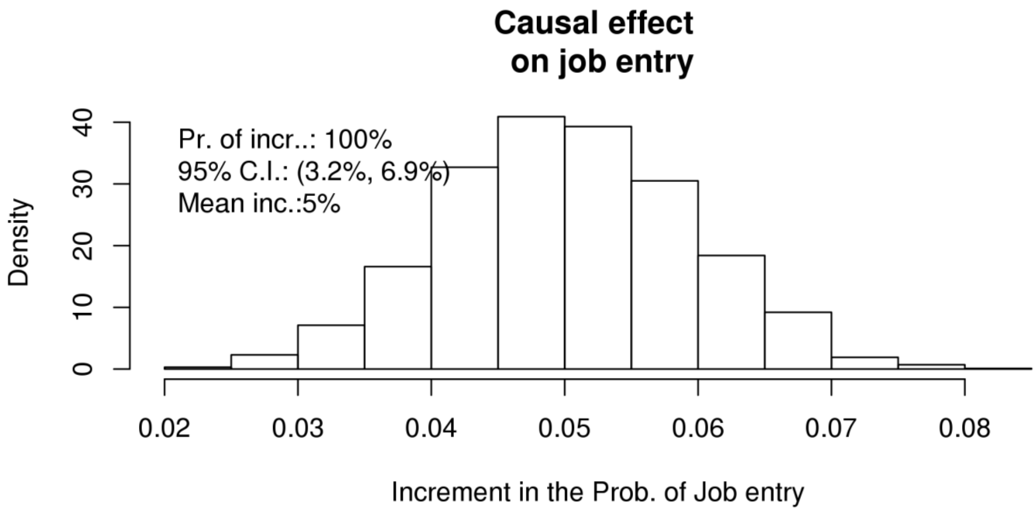
Table 2 Descriptive Statistics: Treatment and Control Groups

	Job entry Mean		Job exit Mean	
Variable	Treated	Control	Treated	Control
Hiquald	4.0539	4.0471	3.9983	3.947
Eth01	1.2403	1.2251	1.2384	1.2537
Appren	3.4707	3.4763	3.5168	3.5295
Uresmc	10.451	10.517	10.518	10.386
stucur	1.8232	1.8102	1.9062	1.9118
Sex	1.6754	1.6609	1.731	1.7268

The analytical 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. Figures 1 and 2 report the estimated distribution of the total causal impact of increases in the minimum wage rate on job exit and job entry using BART model with all workers aged 18-40. The results indicate that treatment has a positive effect both on job entry and job exit that are consistent with the alternative estimations already reported. More specifically, the NMW exerts a positive impact on job entry with a probability of 100%, the mean value of this causal impact is 5% with a 95% confidence interval equal to [3.2%, 6.9%]. For job exit, the effect is positive with 100% probability with mean value equal to 2% and with a 95% confidence interval equal to [1%, 4%]. Note that although the estimated effect on job entry is larger than that for job exit, the overall effect of NMW is ambiguous as there are lot of more employed workers (who are candidates for job exit) than unemployed individuals (candidates for job entry).

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. In Figure 2, we interact gender with the effect ofNMW increases. Again, the previous finding of a greater effect of NMW increases on job exit than on job entry is reproduced. As for gender, it appears to play little role although there is a slightly higher impact for men than female on job entry. In Figure 3, in turn, we consider the interaction with age (expressed in months rather than years). Here, the pattern is quite 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 likely to be vulnerable to NMW increases. Besides, the interactive effect is clearly stronger in job entry. In Figure 4, we consider the interaction with the highest attained qualification. The graphs reveal that this is not a relevant factor to explain differences in the causal impact of NMW either for job entry or job exit. Finally, Figure 5 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 there the business cycle has little bearing on how the minimum wage shapes job exits.

Figure 2 NMW Increases, Job Exit/Entry and Gender

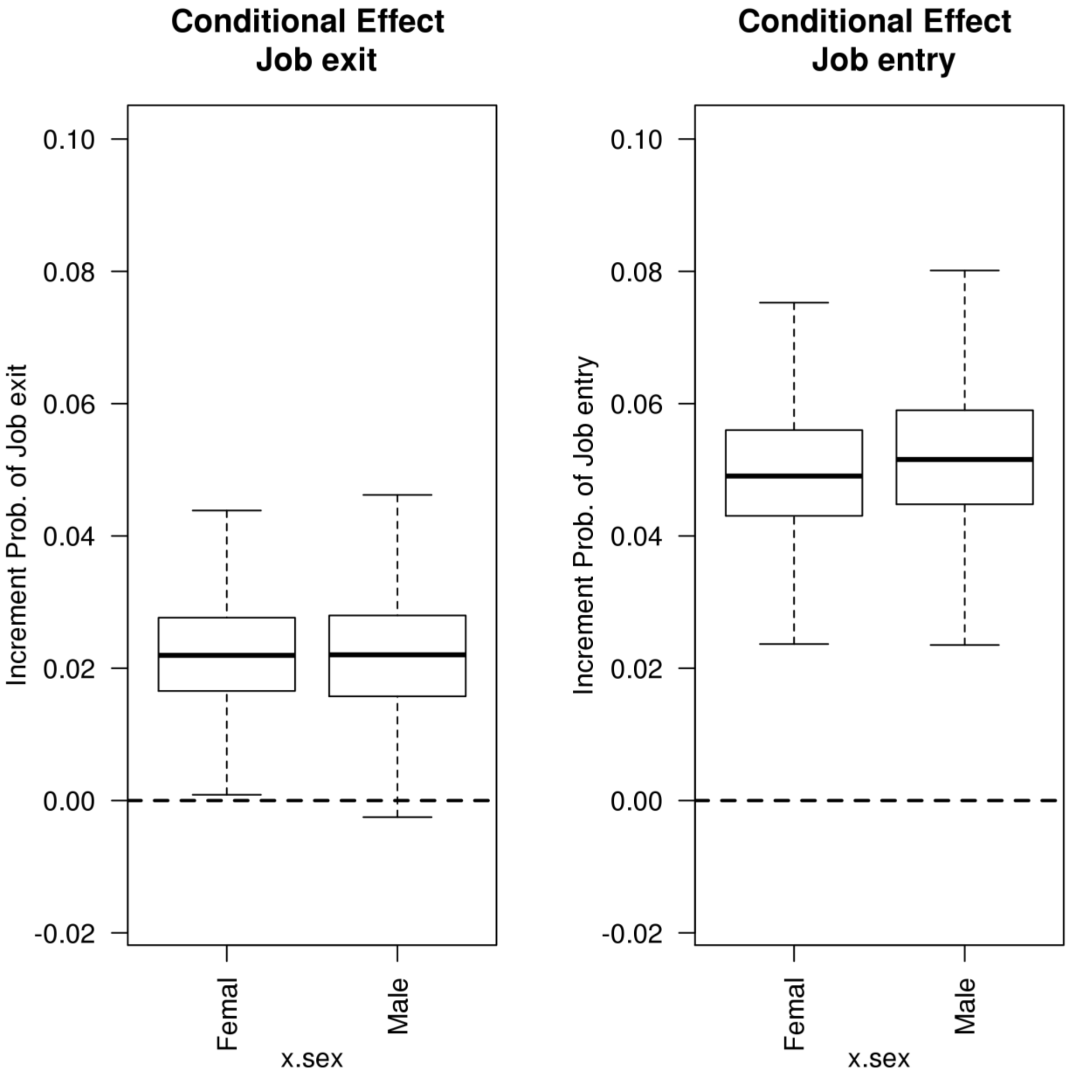


Figure 3 NMW Increases, Job Exit/Entry and Age

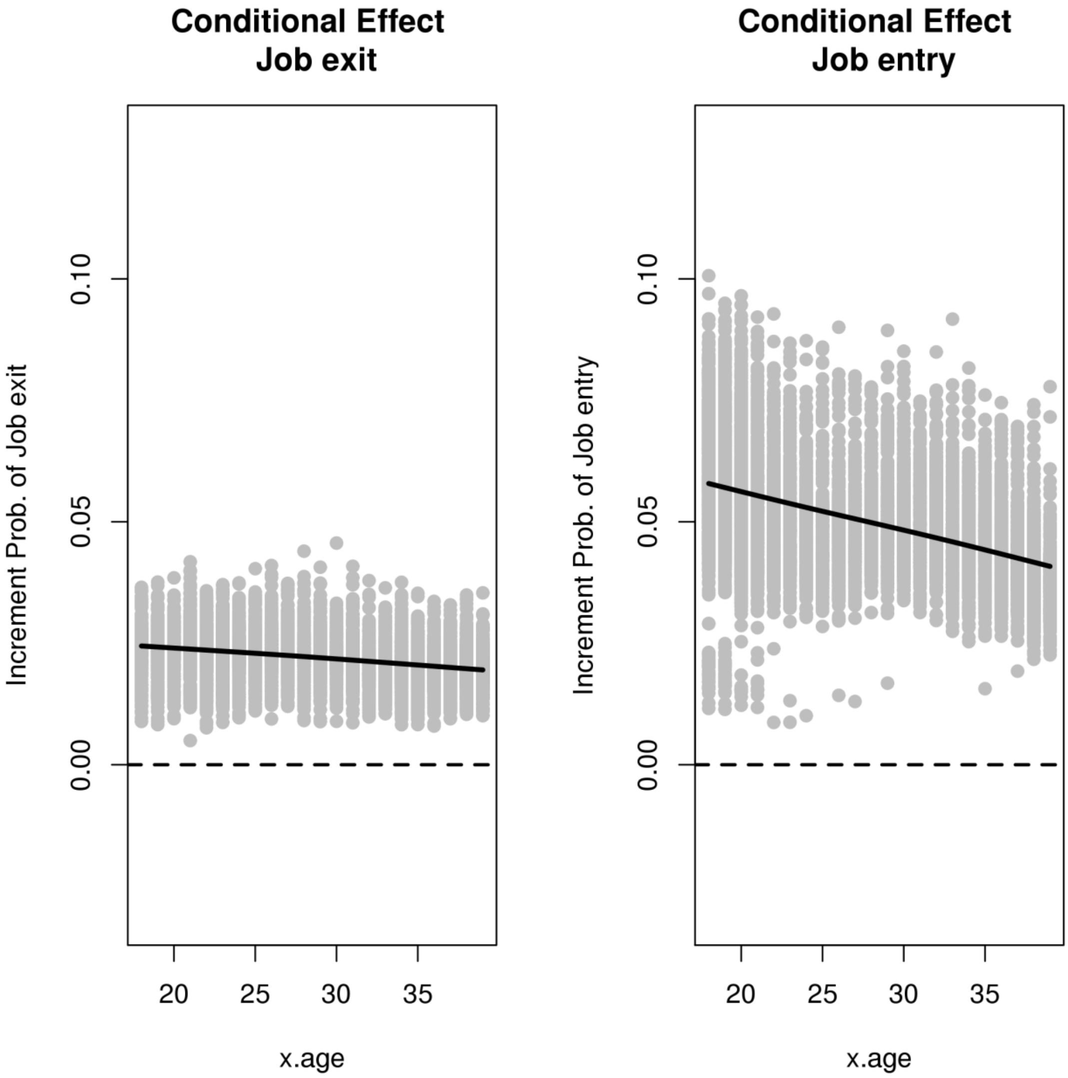
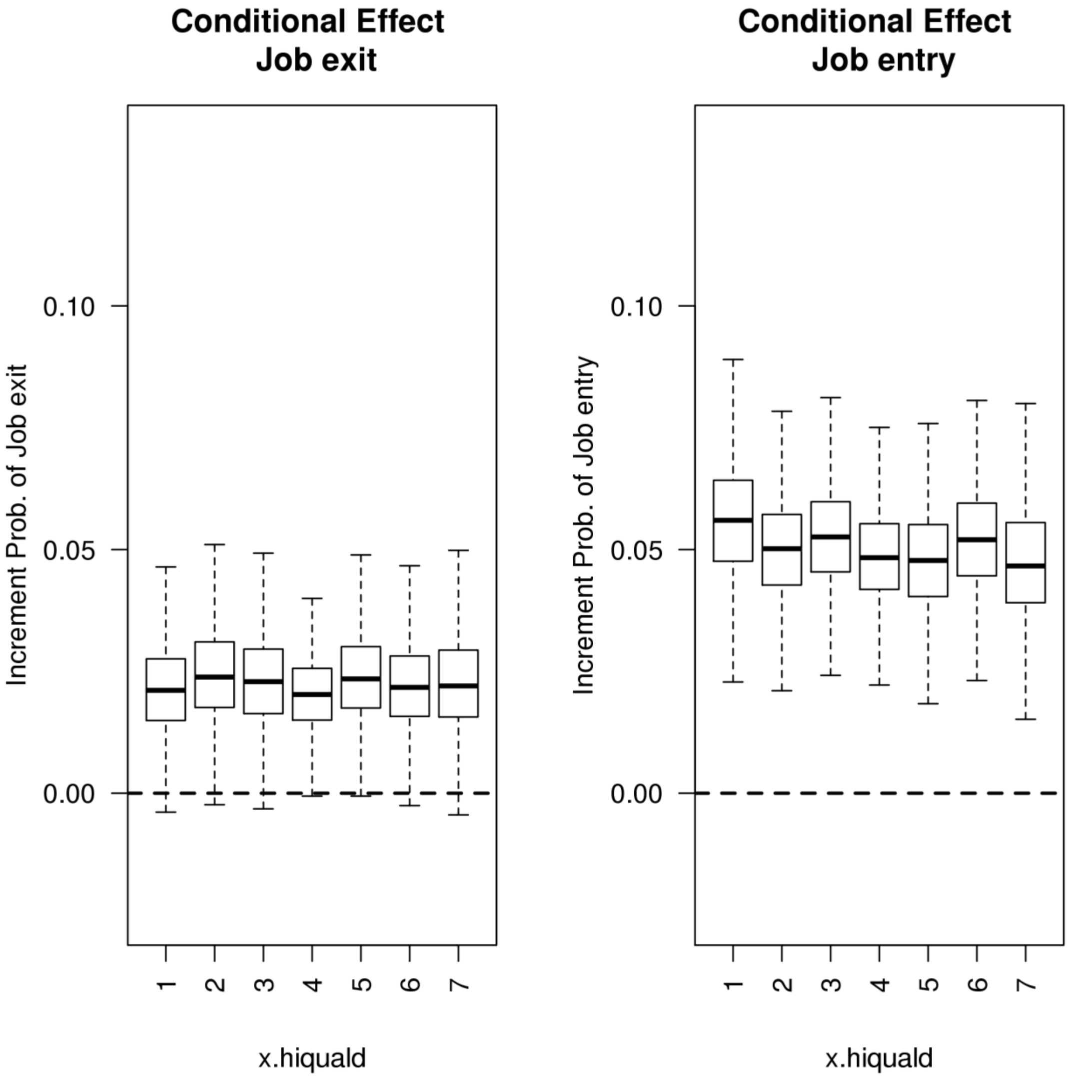
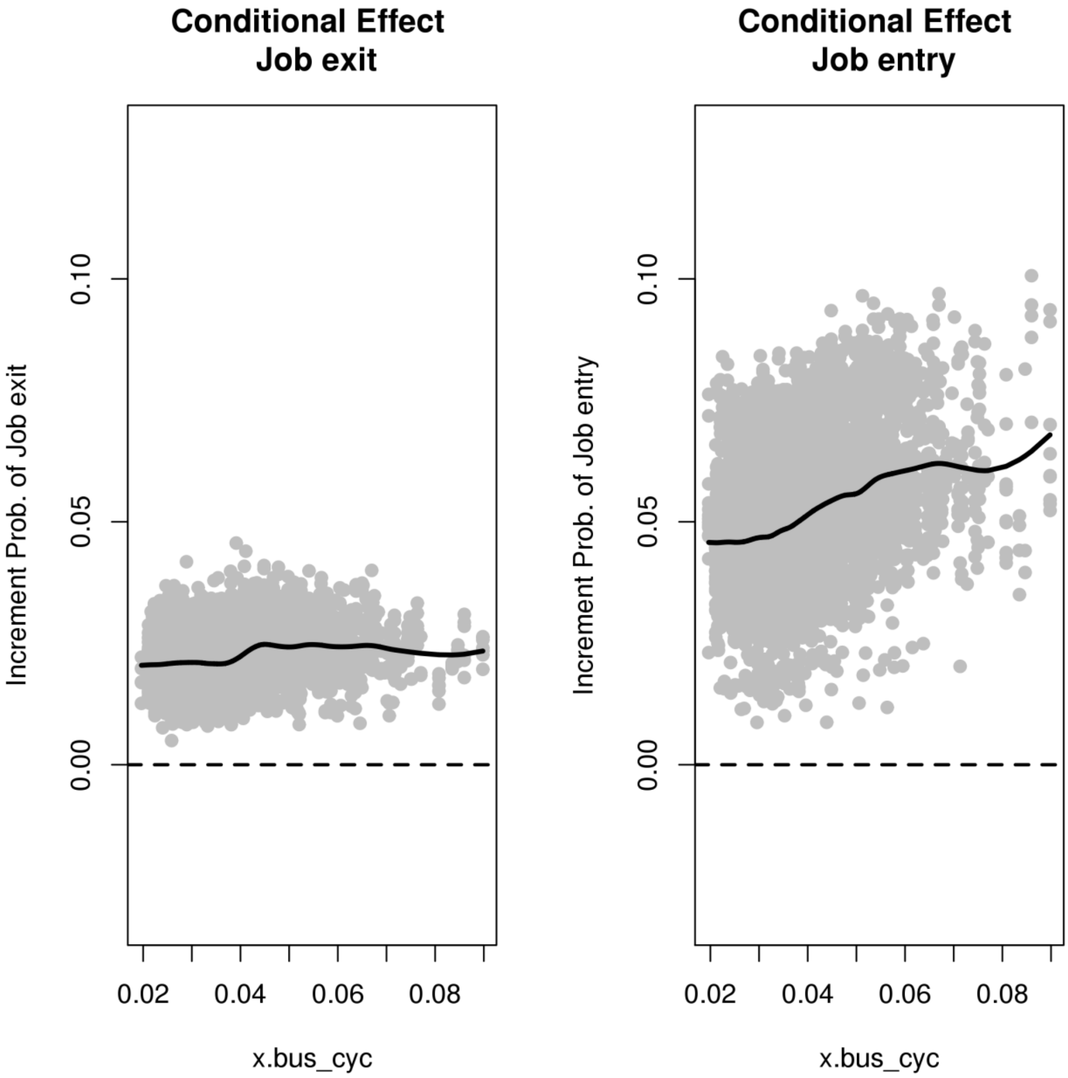


Figure 4 NMW Increases, Job Exit/Entry and Qualifications



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 5 NMW Increases, Job Exit/Entry and Business Cycle



Notes: The horizontal axis measures the regional unemployment rate.

6 Concluding remarks

We estimate the causal impact of the NMW on the probability of job entry and job exit in the UK using a novel and previously little used methodology, Bayesian Additive Regression Trees (BART). An important advantage of this procedure is that it allows for 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 NMW is stronger for job entry than for job exit.

Our paper open new lines of research that can be explored in subsequent works. For example, this fully flexible approach could be adapted to deal with some of the most recent discussion 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." Mimeo.

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.

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.

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.

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.

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.

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.

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

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

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