Lodz Economics Working Papers

Model Specification, Data Selection and the Employment Effect of Minimal Wage

Paweł Strawiński,
Aleksandra Majchrowska

1/2022

FACULTY OF ECONOMICS AND SOCIOLOGY
University of Lodz

Paweł Strawiński, Faculty of Economic Sciences, University of Warsaw, Poland, pstrawinski@wne.uw.edu.pl

Aleksandra Majchrowska, Chair of Macroeconomics, University of Lodz, Poland; aleksandra.majchrowska@uni.lodz.pl

Model specification, data selection, and the employment effect of minimum wage[1]

Abstract

This study aims to show that the effect of minimum wage on employment highly depends on the model specification. To verify this hypothesis, we use a publicly available dataset on employment and minimum wage from the article of Dube et al. (2010) for US counties in 1990–2006. We replicate the minimum wage employment equation using different model specifications and data subsets to deduce the empirical distribution of the minimum wage parameter. In addition, we verify the relationship between the sample size and the significance of the estimated minimum wage parameter.

Our research confirms that the specification of the econometric model determines the results of the minimum wage employment relationship. Different model specifications lead to different results and are incomparable. The results show that allowing for differences in the time trends in different survey areas nullifies the relationship between minimum wage and employment. This was observed only in a few counties, so it was difficult to find a statistically significant relationship at the country level. In addition, our study shows that data selection has no significant impact on the results.

Keywords: data distribution, employment elasticity, minimum wage, model specification, sample size, statistical significance

JEL codes: C51, J21, J31.

1. Introduction

There have been a large number of studies on the impact of minimum wage on employment. Neumark and Shirley (2021), and Wolfson and Belman (2019) present the most recent summaries of evidence from the US. However, studies have also confirmed this issue in other countries (see e.g., Campolieti, 2020 with meta-analysis for Canada, and Dube (2019) with summary of international evidence).

[1] The paper was financed by the National Science Centre Poland, Grant Number: UMO-2017/25/B/HS4/02916

Neumark and Shirley (2021) underline that disagreement on this issue not only exists across individual studies but is also observed in the meta-analyses summarizing the body of the literature.

An important issue often addressed in meta-analyses, which potentially bears on interpreting the evidence, is the publication bias. However, Neumark and Shirley (2021) state that "it is hard to distinguish between publication bias and other sources of patterns in the published evidence consistent with publication bias". For example, meta-analyses like Doucouliagos and Stanley's (2009) argue that if published negative estimates of the minimum wage effects contain large standard errors, then this is evidence of publication bias. Similarly, Andrews and Kasy (2019) identify publication bias in meta-analyses when the distributions of estimates and standard errors are not independent.

Problems in estimating the value of minimum wage parameter also appear because the minimum wage employment relationship is nonlinear, but the linear model often approximates it. Thus, the value of the minimum wage parameter in the linear model will be the function of both the minimum wage and employment level as well as their changes (Christl et al., 2019). The results of previous studies depend on the data that are available. In a high minimum wage setting, it is likely that the estimate will be negative (see Manning, 2016).

In this study, we aim to demonstrate that the effect of the minimum wage on employment highly depends on the model specification. To verify this hypothesis, we use a publicly available dataset on employment and minimum wage from the article of Dube et al. (2010) for US counties during 1990–2006.[2] We perform two empirical tasks. First, we replicate the minimum wage employment equation of Dube et al. (2010) using different model specifications and data subsets to deduce the empirical distribution of the minimum wage parameter. Second, we verify the relationship between sample size and significance of the estimated minimum wage parameter (see Doucouliagos and Stanley, 2009). The remainder of this article is organized as follows. Section 2 briefly describes the data. Section 3 presents the empirical approach and results. Section 4 concludes the study.

2. Data

To show the statistical properties of the minimum wage employment parameter, we use the same dataset as Dube et al. (2010). They used a combination of Quarterly Census of Employment and Wages data, which provide county-level payroll data by industry, and the Bureau of Labor Statistics quarterly data on employment and wages in the restaurant sector. The data concerns US counties. Since the authors aimed to analyze the balanced data panel of 1990–2006, they used 1,381 of the 3,081 counties in the US.

The dependent variable in the model of Dube et al. (2010) is the logarithm of employment in county *i* in year *t*. The main independent variable is the minimum wage level in county *i* in year *t*. They start with the simplest specification, and further include controls for the logarithm of total private sector employment $\ln\left(y_{it}^{TOT}\right)$ and the logarithm of county-level population$\ln(pop_{it})$:

[2] The datasets are provided by Dube et al. (2010) on https://doi.org/10.1162/REST_a_00039.

$$\ln(E_{i,t}) = \alpha + \eta \ln(MW_{i,t}) + \delta \ln(y_{i,t}^{TOT}) + \gamma \ln(pop_{i,t}) + \phi_i + \tau_t + \varepsilon_{it} \qquad (1)$$

where ϕ_i term denotes the county fixed effects; τ_t denotes the time period fixed effects, which are assumed to be constant across the counties.

The authors use a two-step procedure to estimate the minimum wage equation in the original article. In the first step, they regress all variables on county dummies to deduce residuals. In the second step, they analyze the minimum wage employment relationship on the residuals obtained. Therefore, the estimated equation has the following form:

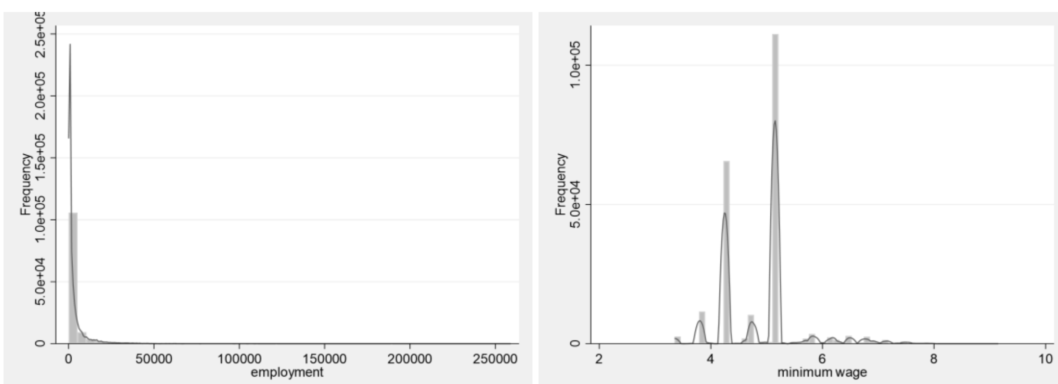
$$\ln(ER_{i,t}) = \alpha + \eta \ln(MWR_{it}) + \delta \ln(yR_{it}^{TOT}) + \gamma \ln(popR_{it}) + \zeta_{it} \qquad (2)$$

where *ER*, *MWR*, *yR*, and *popR* denote residuals from the first step regressions of the original variables on the fixed effects. Thus, the minimum wage employment elasticity estimation does not involve the original data, but residuals from the first step. Mathematically, those procedures are equivalent; however, the transformation of data may influence the true distribution of test statistics.

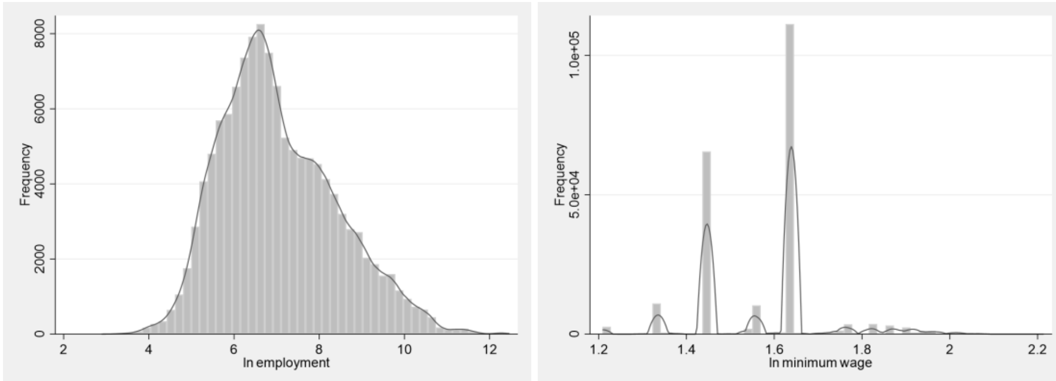
Before replicating the results of Dube et al. (2010), we look more carefully at the statistical data. The distribution of both the level of employment and minimum wage data is far from normal (see Figure 1). Transforming these data from levels into logarithms changes the distribution of data to be more like that of the standard normal distribution. Dube et al. (2010) transformed the data further by regressing all variables on state dummies to deduce residuals. This transformation changes the distribution of the employment data to the normal distribution, with a mean of approximately zero. With the minimum wage variable, the transformations do not significantly alter the shape of the distribution.

Figure 1. Distributions of the employment and minimum wage data in 1,381 counties from 1990 to 2006

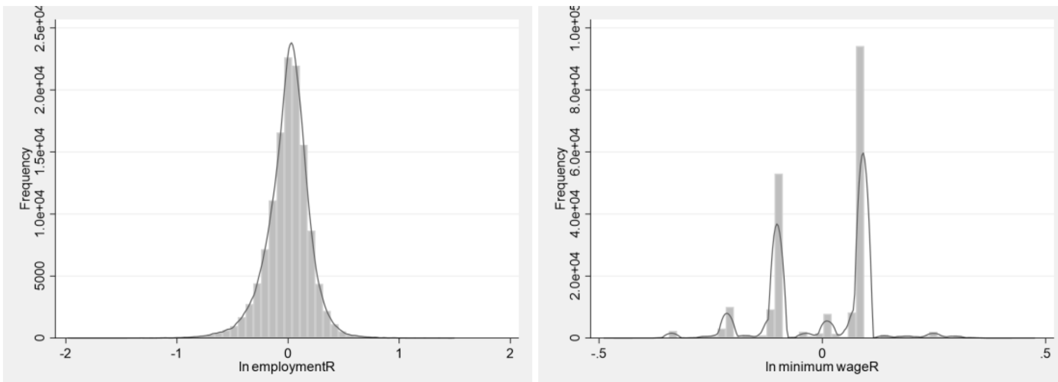
a) Data in levels



b) Data in logarithms



c) Data without fixed effects (residuals from equation 2)



Source: Dube at al. (2011), own calculations.

3. Empirical approach

We begin by replicating the minimum wage employment equation of Dube et al. (2010) using different model specifications and data subsets to deduce the empirical distribution of the minimum wage parameter.

First, we estimate the relationship between minimum wage and employment using the variables in levels. We regress the employment level in county *i* at time *t* on the minimum wage level in county *i* at time *t*, taking the total private sector employment level and county-level population as control variables:

$$E_{i,t} = \alpha + \eta MW_{i,t} + \delta y_{i,t}^{TOT} + \gamma pop_{i,t} + \phi_i + \tau_t + \varepsilon_{it} \tag{3}$$

We then transform all the variables into logarithms, similar to Dube et al. (2010):

$$\ln\left(E_{i,t}\right) = \alpha + \eta \ln\left(MW_{i,t}\right) + \delta \ln\left(y_{i,t}^{TOT}\right) + \gamma \ln\left(pop_{i,t}\right) + \phi_i + \tau_t + \theta_{it} \tag{4}$$

Then, we estimate the minimum wage employment relationship with the residual data, similar to Dube et al. (2010):

$$\ln\left(ER_{i,t}\right) = \alpha + \eta \ln\left(MWR_{it}\right) + \delta \ln\left(yR_{it}^{TOT}\right) + \gamma \ln\left(popR_{it}\right) + \zeta_{it} \tag{5}$$

Following Dube et al. (2010), in the above specifications, we allow the period fixed effects to vary across the nine census divisions τ_{ct} and include state-level linear time trends. Therefore, the model on levels can be written as follows:

$$E_{i,t} = \alpha + \eta MW_{i,t} + \delta y_{i,t}^{TOT} + \gamma pop_{i,t} + \phi_i + \tau_{ct} + \vartheta_s I_s * t + \varepsilon_{it} \quad (6)$$

where I_s is a dummy for state s, and ϑ_s is a state-specific trend. Similar modifications were made to equation (4) on the variables in logarithms and to equation (5) on residuals:

$$\ln{(E_{i,t)}} = \alpha + \eta \ln\left(MW_{i,t}\right) + \delta \ln\left(y_{i,t}^{TOT}\right) + \gamma \ln\left(pop_{i,t}\right) + \phi_i + \tau_{ct} + \vartheta_s I_s * t + \theta_{it} \quad (7)$$

$$\ln{(ER_{i,t)}} = \alpha + \eta \ln(MWR_{it}) + \delta \ln\left(yR_{it}^{TOT}\right) + \gamma \ln(popR_{it}) + \tau_{ct} + \vartheta_s I_s * t + \zeta_{it} \quad (8)$$

To deduce the entire distribution of the minimum wage parameter, we use the panel of observations for the above six models (equations 3–8), in each case, dropping the observations for two states. Finally, we obtained 1,275 estimates for each model; thus, we were able to deduce the distribution of the parameter of employment with respect to the minimum wage.

Estimating the distribution of minimum wage–employment relationship

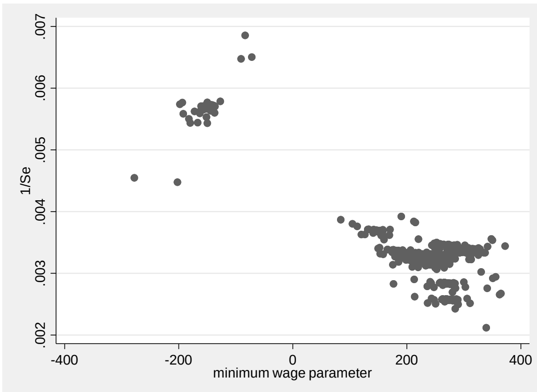
Table 1 in the Appendix presents the results of the estimated distribution of the parameter between the minimum wage and employment. While conducting the estimation with the variables in levels (equation 3), all estimated parameters are statistically insignificant and most are positive. When the variables are transformed into logarithms (equation 4), the results change. The estimated parameters remain statistically insignificant, but most of the estimates are negative. In the model estimated on residual variables (equation 5) the distribution of the parameter change further; 75% of the estimates are statistically significant (at the 10% significance level) and negative (see Table 1). When the period fixed effects are varied and state-level linear time trends included (equations 6–8), almost all estimated parameters are insignificant.

Separating data selection bias from minimum wage employment relationship

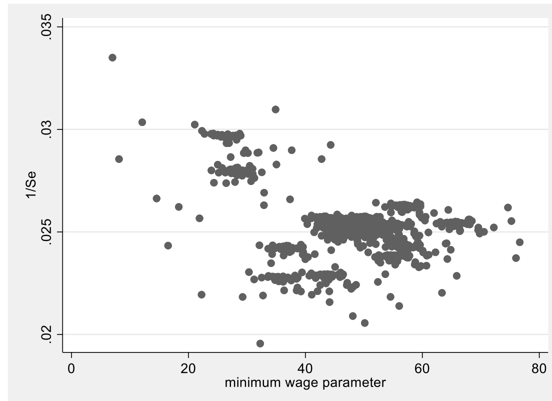
Next, we investigate the existence of data selection bias. To address this problem, we follow Doucouliagos and Stanley (2009) and adapt meta-regression analysis. We begin by analyzing the scatter plots of precision versus effect: inverse of the standard error versus estimated value of the minimum wage parameter in our case.

Figure 2. Correlation between the estimated value of the minimum wage parameter (horizontal axis) and the inverse of the standard error (vertical axis)

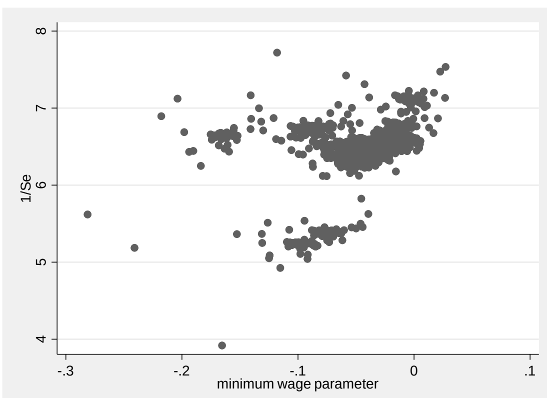
Equation (3)



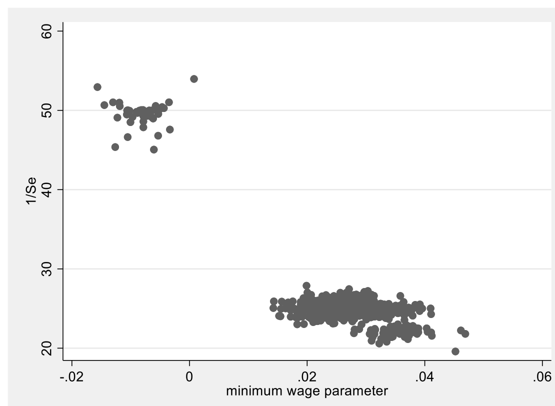
Equation (6)



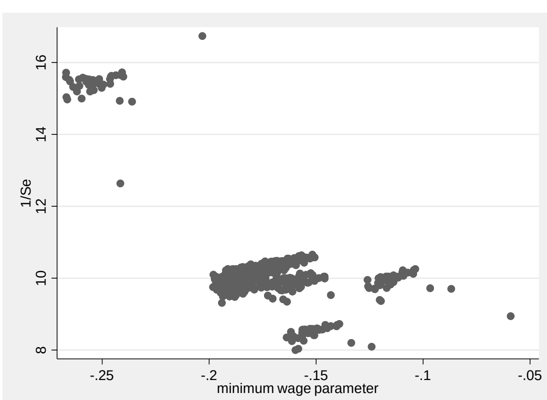
Equation (4)



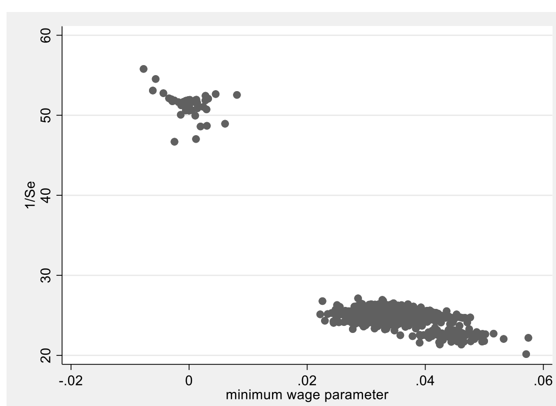
Equation (7)



Equation (4)



Equation (7)



Source: own estimations.

To formally test for an existence of empirical effect beyond eventual bias arising from sample selection, we adapt and used meta-significance testing method (see Doucouliagos and Stanley, 2009):

$$E(\ln|t|) = \alpha_0 + \alpha_1 df \quad (9)$$

where *t* is the t-value of estimated parameter and *df* is the number of degrees of freedom in the model. Positive and significant α_1 provides evidence of true empirical effect.

The results (see Table 2 in Appendix) show that sample size has a significant impact on the importance of the estimated parameter in the estimation of variables in levels (equation 3). The bigger the sample, the higher is the t-value of the estimated parameter. Although the relationship is significant from a statistical viewpoint, it is quantitatively insignificant.

The relationship between degree of freedom and estimated t-value is insignificant in models with variables in logarithms and logarithms of residuals (equations 4–5). For equations 6–8, the estimated relationship is significant, but the value of the parameter is virtually zero.

4. Conclusions

Our research shows that the results of the minimum wage employment relationship highly depend on the specification of the econometric model. Transformation of the data has a significant impact on the distribution of estimates. Therefore, while conducting analyses, especially meta-analyses, the functional form of the model should be selected carefully. Different model specifications lead to different results and are incomparable. Inappropriate selection of models compared (with different functional forms) may lead to misleading conclusions.

The results show that allowing for differences in time trends in different survey areas nullifies the relationship between minimum wage and employment. The relationship is area-specific (i.e., county- or state-specific) and derived from macroeconomic trends. This was observed in a few counties; however, it was difficult to find a statistically significant relationship at the country level.

In models estimated on logarithmic transformed variables, data selection had no significant impact on the results. Data selection was also not an issue for the model with census area-specific time trends. The estimated elasticity for the latter model suggests that the working population should be four to five times larger to increase the significance of the minimum wage employment relationship by 1%.

References:

Andrews, I., Kasy, M. (2019) Identification and Correction for Publication Bias, American Economic Review, 109(8), 2766-2794.

Brown, A. J., C. Merkl, and D. J. Snower (2014), The minimum wage from a two-sided perspective, Economics Letters, 124 (3), 389-391.

Campolieti, M. (2020) Does an Increase in the Minimum Wage Decrease Employment? A Meta-Analysis of Canadian Studies. Forthcoming in Canadian Public Policy, 46(4), 531-564.

Christl, M., Köppl-Turyna, M., Kucsera, D. (2019) Employment Effects of Minimum Wages, ifo DICE Report, 16(04), 01-08.

Doucouliagos, H., Stanley, T.D. (2009) Publication Selection Bias in Minimum-Wage Research? A Meta-Regression Analysis, British Journal of Industrial Relations, 47(2), 406-28.

Dube, A. (2019) Impacts of Minimum Wages: Review of the International Evidence, https://www.gov.uk/government/publications/review-of-the-international-evidence-on-the-impacts-of-minimum-wages/review-of-the-international-evidence-on-the-impacts-of-minimum-wages-terms-of-reference.

Dube, A., Lester, T, Reich, M. (2010) Minimum wage effects across state borders: estimates using contiguous counties, The Review of Economics and Statistics, vol. 92/4, pp. 945-964.

Dube, A, Lester, T., Reich, M. (2011) Replication data for: Minimum Wage Effects across State Borders: Estimates Using Contiguous Counties, https://doi.org/10.7910/DVN/L4DUZ7, Harvard Dataverse, V2.

Neumark, D., Shirley, P. (2021) Myth or measurement: What does the new minimum wage research say about minimum wages and job loss in the United States?, National Bureau of Economic Research.

Wolfson, P., Belman, D. (2019) 15 Years of Research on US Employment and the Minimum Wage, LABOUR, 33(4), 488-506.

Table 1. The estimated parameters of the minimum wage–employment relationship in equations (3–8)

	Equation					
Percentiles	(3)	(4)	(5)	(6)	(7)	(8)
1	-158.135 (-0.901)	-0.171 (-1.136)	-0.258*** (-3.983)	23.922 (0.668)	-0.010 (-0.487)	-0.001 (-0.068)
5	141.528 (0.520)	-0.110 (0.688)	-0.194* (-1.941)	28.278 (0.786)	0.018 (0.432)	0.025 (0.634)
10	183.360 (0.609)	-0.091 (-0.559)	-0.188* (-1.897)	35.798 (0.861)	0.021 (0.523)	0.029 (0.718)
25	219.959 (0.711)	-0.062 (-0.388)	-0.183* (-1.851)	45.142 (1.136)	0.024 (0.600)	0.032 (0.797)
50	233.673 (0.757)	-0.045 (-0.290)	-0.177* (-1.807)	47.986 (1.217)	0.027 (0.680)	0.035 (0.869)
75	251.782 (0.798)	-0.032 (-0.208)	-0.168* (-1.734)	51.361 (1.291)	0.030) (0.747)	0.037 (0.928)
90	280.688 (0.933)	-0.016 (-0.104)	-0.158 (-1.564)	56.251 (1.399)	0.034 (0.802)	0.041 (0.992)
95	302.079 (0.999)	-0.007 (-0.048)	-0.147 (-1.272)	59.680 (1.512)	0.036 (0.849)	0.045 (1.046)
99	330.351 (1.081)	0.006 (0.044)	-0.112 (-1.118)	67.889* (1.728)	0.039 (0.949)	0.048 (1.128)
Number of observations	1,275	1,275	1,275	1,275	1,275	1,275
Mean value	221.669 (0.705)	-0.051 (-0.326)	-0.176* (-1.831)	47.424 (1.195)	0.026 (0.639)	0.034 (0.837)
Standard deviation	83.846 (0.338)	0.035 (0.222)	0.022 (0.455)	8.665 (0.211)	0.008 (0.235)	0.008 (0.198)

*Estimated parameter (t-statistic).

Source: own calculations.

Table 2. Estimated parameter of equation (9)

	Model specification					
	(3)	(4)	(5)	(6)	(7)	(8)
df	0.00001*** (5.37)	-0.00001 (-1.18)	-0.000003 (-1.12)	0.00003*** (0.000)	0.00003*** (8.97)	0.00009*** (10.94)
Number of observations	1,275	1,275	1,275	1,275	1,275	1,275
Adj. R^2	0.021	0.0003	0.0002	0.080	0.059	0.085

Estimated parameter (t-statistic)

Source: own calculations.