Gelbach in Logit:
A Covariate Decomposition for the Logit Model applied to the Minimum Wage's Heterogeneous Impact

Miguel Salema

Dissertation written under the supervision of professor Pedro Raposo

Dissertation submitted in partial fulfillment of requirements for the MSc in Economics with Specialization in Macroeconomic Policy, at the Universidade Católica Portuguesa, December, 2022.

Gelbach in Logit: A Covariate Decomposition for the Logit Model applied to the Minimum Wage's Heterogeneous Impact

Miguel Salema

Supervisor: Pedro Raposo

December, 2022

Abstract

Sequentially adding control variables to a regression to investigate their effect on a structural parameter is econometrically meaningless when controls are intercorrelated, as the order in which control variables are added will influence how the structural parameters change. As a solution, I develop a novel order-invariant conditional decomposition for the logit model. Furthermore, this logit decomposition can explain which variables are responsible for the heterogeneous treatment effect on the treated. I illustrate the utility of the decomposition with an application. Using a natural experiment to estimate the displacement effects of the minimum wage in Portugal, I find its effects to be heterogeneous. Moreover, by using the decomposition, I find that the heterogeneous impacts are 65% explained by firms, 28% by the worker, and 7% by tenure; implying that the primary determinant of the minimum wage effect on workers' displacement is the firm they work for.

Keywords:

Gelbach Decomposition; Logit Covariate Decomposition; Logit Two-Way Fixed Effects; Minimum Wage; AKM model; Separations; Triple Difference Estimator; Natural Experiment.

Gelbach em Logit: Decomposição de Covariáveis no Modelo Logit aplicada ao Impacto Heterogéneo do Salário Mínimo

Miguel Salema

Supervisor: Pedro Raposo

Dezembro de 2022

Resumo

A adição sequencial de variáveis de controlo numa regressão base, tanto com a intenção de quantificar os efeitos dos controlos num determinado parâmetro como para fins de robustez, não tem significado econométrico quando as variáveis de controlo estão correlacionadas. Desenvolvo uma nova decomposição condicional no modelo logit, que não varia consoante a ordem em que as variáveis são adicionadas. Além disso, esta decomposição logit pode explicar quais as variáveis responsáveis pelo efeito heterogéneo que um tratamento tem no grupo tratado. Aplico a decomposição logit a uma experiência natural no salário mínimo português. Descubro que as diferenças no impacto do SM nas separações são em 65% explicadas pelas empresas, em 28% pelo trabalhador e em 7% pela antiguidade, tornando a empresa no principal determinante do impacto do salário mínimo sobre um trabalhador.

Palavras-chave:
Decomposição de Gelbach; Decomposição por Covariáveis em Logit; Logit com 2 Efeitos Fixos; Salário Mínimo; Experiência Natural; Modelo AKM; Separações de Empresas; Estimador de Diferença Tripla.

Acknowledgments

Firstly, I'm extremely grateful to my supervisor, Pedro Raposo. When I decided I wanted to do a thesis about econometric theory, not only didn't he attempt to persuade me into doing something easier, he pushed me further and had monk-like patience with all the complications my stubbornness created. For that, and much more, I express my deepest gratitude. Secondly, I'm thankful to Jaime Montana for teaching me how to write, a skill I had until now undervalued. Thirdly, many thanks to all my colleagues at Prosper, who gave me excellent advices. Finally, I could not have undertaken this journey without the support of my family, friends, and girlfriend, Mafalda, who read this thesis more times than I did.

Contents

Divide each difficulty into as many parts as is feasible and necessary to resolve it.

René Descartes,
Discourse on the Method (1637)

1 Introduction

Several econometric problems require finding how each variable from a set of covariates contributes to the change of some coefficients of interest in a regression. The purpose is either to test the robustness of the coefficients of interest by measuring their stability across specifications or to account for the individual impact of each covariate on those coefficients of interest. There were no tools to perform this analysis in a logit regression.

This thesis' main contribution is a novel logit conditional covariate decomposition that splits the contribution of each control variable in explaining the effects of structural variables.[1] The decomposition has 2 parts, as stated in equation (1), with two distinct interpretations. The first part, the confounding effects, decomposes the effects resulting from the correlation between the control variables and the structural variables. They are interpretable as the effect that each omitted variable has on the structural variables' explanatory power. The second part, the rescaling effects, decomposes the correlation between the control variables and the structural variables that *only* exists conditional on the state of the dependent variable. The decomposition of the rescaling has a particular interpretation. It's a division, between the control variables, of the heterogeneity of treatment effects on the treated.

The logit decomposition can answer many questions involving binary dependent variables.[2] For example, "Why are women more likely to go to college?" With the decomposition of the confounding effects, we can divide the contribution of several variables, like wealth, high-school courses, professional experience, and psychological tests, and find their individual contribution in making women more college-prone. The rescaling effects answer a different question. Since it's expected that sex won't affect the probability of going to college in the same way for all individuals, with this decomposition, we can find the contribution of each variable to this heterogeneity. This

[1] See Fortin, Lemieux, and Firpo (2011): there isn't a Gelbach covariate decomposition for the logit model nor any other GLM.

[2] Non-linear models are more appropriate to study a binary dependent variable than linear probability models. Both because of heteroskedasticity and, most importantly, due to bad fit. It's unconvincing to assume that the effects of a given set of variables are linear throughout the assumed probability distribution. Moreover, and because of being a bad fit, the predicted probabilities in a linear probability model will go beyond the limits of 0 and 1 (Wooldridge 2010).

allows us to find which variables make the effect of sex_i materializing in college admissions. This exercise can be done with any binary problem, like firms that grow/don't grow, consumers who buy/don't buy, people who migrate/don't migrate, or a tennis player winning/losing.

Logit omitted variable bias:

$$\widehat{logit\ bias}_{x_1} = \hat{\boldsymbol{\beta}}_{\mathbf{1}}^{base} - \hat{\boldsymbol{\beta}}_{\mathbf{1}}^{full} = \underbrace{\text{confounding effects}}_{\substack{\text{Decomposable using Gelbach} \\ \text{between variables } \mathbf{Z}}} + \underbrace{\text{rescaling effects}}_{\text{Unexplained with Gelbach}}, \tag{1}$$

where $\hat{\boldsymbol{\beta}}_{\mathbf{1}}^{base}$ is a matrix with coefficients from a logit regression of $\mathbf{X}$ on the binary variable $\mathbf{Y}$, and $\hat{\boldsymbol{\beta}}_{\mathbf{1}}^{full}$ continue to be the coefficients of $\mathbf{X}$, but on a logit regression with $\mathbf{X}$ and $\mathbf{Z}$ (other covariates) on $\mathbf{Y}$.

The confounding effects in logit are equivalent to the full omitted variable bias (OVB) of linear models, where OVB stems from a correlation between **X** and **Z**.[3] Thus, to decompose the confounding effects I use the Gelbach (2009) decomposition for linear models. Before Gelbach (2009), whenever researchers wished to find how each **Z** variable affected the sensitivity of some structural parameters, $\boldsymbol{\beta}_1$, in a linear model, they would sequentially add them one by one to a regression with only the structural variables **X** and register the change in $\boldsymbol{\beta}_1$. However, such practice lacks econometric meaning. The covariances within **Z** make sequential addition (or subtraction) order-variant. Meaning that the order chosen by the econometrician will shape the results. Gelbach (2009) solves this problem for linear models by building a purposefully endogenous base model with only **X**. The endogeneity is then divided amongst **Z** by using Golberger (1991) omitted variable bias (OVB) formula. This simple method is capable of generating an order-invariant meaningful conditional decomposition of any parameter in a linear regression.

Nonetheless, Gelbach's work cannot be directly applied to the logit model due to equation's (1) second part, the rescaling effects. The rescaling effects are the consequences of a necessary feature of the logit model's estimation (also present in the probit model, see Cramer (2007)). In linear models, the dependent variable is given and the residuals are a result of the estimation. However, in logit models it's reversed: the distribution of the residuals is assumed (generally as a standard logistic distribution) and the *log-odds* estimated (*log-odds* are the left-hand side in models (4) and (3)). That is because the maximum likelihood estimator cannot solve an equation for two residuals. As a consequence, the "true" residuals of a logit model must be coerced to a standard logit distribution and fixed as such. By doing so, all the parameters of the logit model are rescaled.

Because the distribution of the logit residuals is fixed, whenever a variable is added the whole

[3] Besides $corr(\mathbf{X}, \mathbf{Z}) \neq 0$, for OVB to exist as confounding effects we further need an existing correlation of both **X** and **Z** with **Y**

model must be rescaled, since the residuals cannot change. This feature creates a seemingly bizarre consequence: logit parameters may be biased by totally uncorrelated variables. To decompose these rescaling effects, I characterize the rescaling part from an OVB standpoint, using L. Lee (1980) logit OVB formula. The procedure consists in computing the residuals from the regressions of each **Z** on **X** and then computing the correlation of those residuals with **X** conditional on the state of **Y** being 1. However, the main caveat of this logit decomposition is the assumption that the residuals of regressing each variable from **Z** on **X** follow a normal distribution conditional on **X** and the dependent variable. I build a simulation that demonstrates this problem. Furthermore, I explain in this thesis why the rescaling effects can decompose the heterogeneity of the effects of treatment on the treated.

Thus, for the confounding effects, this logit covariate decomposition uses Gelbach (2009) and, for the rescaling effects, employs a similar logic to describe the effects as OVB and divide it among **Z**, but with a different bias formula. I build an R package called "Decomp" to implement this procedure.

To demonstrate the proposed method, I study the effect of an increase in the minimum wage (MW) on the likelihood of a MW worker being displaced. To identify the MW shock I use a natural experiment from Portugal. In 1974, when the national MW was first introduced in Portugal, it did not cover all ages. Until the end of 1986, 18- and 19-year-old workers were granted only 75% of the MW. That changed in 1987 when the teen MW was eliminated, leading to different MW increases *within* MW workers. MW workers that were teens had a nominal MW hike of 36.25% while non-teen workers had an increase of 2.19%. To study the MW natural experiment I build a logit model, with a triple difference methodology and worker and firm FE, where the dependent variable is separations.[4]

I find that the MW hike increased the probability of separation of teen MW workers (the ones directly affected) by 6 percentage points. This is the average treatment effect on the treated (ATT), and, generally, it's the focus of the MW literature; the most common is the employment-MW elasticity (Harasztosi and Lindner (2019) or Jardim et al. (2022)). Nonetheless, because the ATT is an average, it's normal that behind it lies heterogeneity in the treatment effect on the treated, i.e., not all teen MW workers are affected equally: *some* MW workers lose their jobs because of the MW, while others don't. What factors explain the different responses *within* directly affected workers? I can answer it by using the second part of the logit decomposition, the rescaling effects. I conclude that the heterogeneity of the impact of treatment (teen MW) on the treated (teen MW workers) is 65% explained by firms, 28% by workers FE, and 7% by tenure. Thus, the firm where a worker is employed is the most important factor in determining whether or not he will be displaced because

[4] A separation is defined as an employed worker not being employed at the same firm in the following year.

of the MW increase. It's important to know which factors are more likely to exacerbate the negative MW effects in order to design good policies, which avoids them.

This thesis makes the following contributions. Firstly, I provide a novel decomposition for the logit model which decomposes the impact of control variables on structural parameters. Secondly, I develop a method that divides the importance of control variables in explaining the heterogeneous impacts of treatment on the treated. Thirdly, I build an R package called "Decomp" to implement the developed procedure. Lastly, I apply these new econometric methods to a Portuguese natural experiment of the MW; and conclude that the firm where a worker is employed is the most important factor in deciding whether or not he separates in response to a MW hike.

2 Literature Review

Before Gelbach (2009), anytime researchers wanted to find out how each variable influenced the sensitivity of some structural parameters, they would sequentially add those variables, one by one, to a regression, and note the impact of each new variable. For example, on the topic of the racial wage gap, Hellerstein and Neumark (2008), Charles and Guryan (2008), and Bound and Freeman (1992) sequentially add control variables, like differences in education and language skills, to a linear regression with a race dummy. They claim that the change in the race parameter provoked by the addition of each control is the part of the wage gap explained by that given covariate.

However, as described in the introduction, adding variables sequentially is not an econometrically meaningful procedure. The logit cumulative distribution is not an addictively separable function. Thus the development of the Gelbach decomposition for linear models. Gelbach (2009) decomposition solves this problem by diving the omitted variable bias steaming from the absence of those covariates, which is order-invariant.

Nowadays, the Gelbach decomposition is widely used: whether by Carruthers and Wanamaker (2017) on what explains the racial wage gap; Card, Cardoso, and Kline (2015) for what explains the gender wage gap; or Buckles and Hungerman (2013) for the association between season of birth and several outcomes.

Nonetheless, the Gelbach decomposition uses the OVB formula of Golberger (1991), which is not fully applicable to the logit model, and, therefore, the decomposition isn't either. This is because, in addition to variable confounding effects, the logit model has rescaling effects.

The rescaling effects were deeply explored for the first time in Amemiya (1985) textbook, although many times referred to as unobserved heterogeneity. Mood (2009) does a concise literature review discussing its consequences, consequences that go beyond tarnishing decompositions. The rescaling effects contaminate comparisons across models with the same dependent variable and also comparisons between groups in the same model, like men and women or skilled and non-skilled workers. Some methods were developed to fight these issues. To compare between model's coefficients, Karlson, Holm, and Breen (2012) develop a method that improves the method developed by Winship and Mare (1984). The group comparison problem can be reduced by applying Williams (2009) procedure, an improvement of Allison (1999) method, or by taking average partial effects like Wooldridge (2010) suggests (a solution I employ in Section A.3.6 to compare minimum-wage workers with non-minimum-wage workers).

Thus, to develop a logit decomposition, I need an omitted variable bias (OVB) formula for logit that reflects both the confounding and the rescaling effects. Equation (2) is the OVB formula most widely used by the literature, seen in prominent papers like Winship and Mare (1984), Cramer (2007), Allison (1999), or Mood (2009).[5]

$$\hat{\beta}_1^b = \left(\hat{\beta}_1^f + \underbrace{\sum_{z=1}^{p} \left[\hat{\Gamma}_1^z \hat{\beta}_2^z \right]}_{\substack{\text{Confounding} \\ \text{bias}}} \right) \underbrace{\frac{\sqrt{Var(\hat{u}^b)}}{\sqrt{Var(\hat{u}^f)}}}_{\substack{\text{Rescaling} \\ \text{bias}}}, \tag{2}$$

where the β's and Γ's are from the models (3) and (4), and $\hat{u}^f$ and $\hat{u}^b$ are the estimated residuals of the full and base models, respectively.

This strand of literature is useful to transmit the intuition behind the rescaling effects. In a logit model, on top of the familiar confounding effects from the linear model, all coefficients are also rescaled with the standard errors ratio of the model's residuals. Thus, excluding a covariate correlated with the dependent variable will increase the variance of the error, therefore also increasing the model's coefficients.

However, I cannot use this paradigm to develop a logit decomposition, since equation (2) is not useful to analytically characterize logit OVB. For equation (2) to hold, both the full model residuals u^f and base model residuals u^b need to be perfectly logistically distributed. In reality, this assumption is very unlikely to hold. Therefore, coercing the true residuals to an assumed standard logistic distribution makes the whole distribution of the residuals change, not just the variance. Hence my wielding of a forgotten literature strand developed in L. Lee (1982) and L. Lee (1980), which uses Bayes' Theorem to analytically characterize the bias and overcome the aforementioned difficulties.

[5]See Section A.5.3 for a brief proof of equation (2) and Mood (2017) for a deeper explanation through various lenses.

3 The Logit Decomposition

As noted in Sections 1 and 2, I apply a logit omitted variable bias formula developed by L. Lee (1982) to achieve an adaptation of the Gelbach (2009) decomposition for the logit model. In this section, I go through the econometrics behind the logit decomposition.[6] Consider the following models:

The general full logit model:

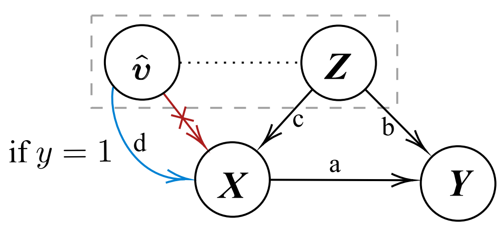
$$ln\left(\frac{P^f(y_i=1|x_i,\mathbf{Z}_i)}{P^f(y_i=0|x_i,\mathbf{Z}_i)}\right) = \hat{\beta}_0^f + \hat{\beta}_1^f x_i + \mathbf{Z}_i\hat{\boldsymbol{\beta}}_2 \tag{3}$$

The general base logit model:

$$ln\left(\frac{P^b(y_i=1|x_i)}{P^b(y_i=0|x_i)}\right) = \hat{\beta}_0^b + \hat{\beta}_1^b x_i \tag{4}$$

where i are observations; y_i is the binary dependent variable; x_i is the structural variable; $\mathbf{Z}_i$ is a $1\times p$ matrix containing a set of control variables z^1, ..., z^p for each i; P are the estimated probabilities; $\hat{u}_i$ are the estimated residuals (not shown in the equations).

Diagram 1: Omitted Variable Bias in the Logit Model



Notes: This DAG diagram illustrates the omitted variable bias in the logit model. **Y** is the binary dependent variable, **X** are the structural variables, **Z** are the omitted variables, and $\hat{\boldsymbol{\upsilon}}$ are the residuals from a linear regression of each **Z** on all **X** (**X** as dependent variables). Thus, $\hat{\boldsymbol{\upsilon}}$ are the parts of **Z** that are uncorrelated with **X** but may be correlated with **X**, conditional on $y=1$. The reader is used to the right-hand-side section of the Diagram, showing that if **Z** is correlated with both **Y** and **X** the analysis is biased. However, the logit model is further biased by the blue curved line, which is a correlation between **X** and $\hat{\boldsymbol{\upsilon}}$, conditional on the state of **Y** being 1.

From the base model (4) to the full model (3) there are 2 sources of bias ($\widehat{bias}_{\beta_1} = \hat{\beta}_1^b - \hat{\beta}_1^f$), also shown in the DAG Diagram 1:[7]

1. A *confounding effect* between x_i and the omitted $\mathbf{Z}_i$ if $\mathbf{Z}_i$ is also correlated with y_i. In Diagram 1, the confounding effects are the product of b and c. They can be decomposed by applying Gelbach (2009);

[6]Although this section conveys the case for a sole main variable x_i, all the deductions remain valid if x is a set of variables $\mathbf{X}_i$.

[7]See Glymour, Pearl, and Jewell (2016) for more details on the notation of DAGs.

2. A *rescaling effect* due to the correlation conditional on the state of y between x and $\hat{\boldsymbol{\upsilon}}$, where $\hat{\boldsymbol{\upsilon}}$ are the estimated residuals from regressions (6), and if $\hat{\boldsymbol{\upsilon}}$ is also correlated with y_i. In Diagram 1, the rescaling effects are the product of d and b, because $\hat{\boldsymbol{\upsilon}}$ and y_i are always correlated in the same way as $\mathbf{Z}_i$ and y_i. For example, both sex and IQ predict college admission, since women are more college-prone. Moreover, sex and IQ are uncorrelated variables. Accordingly, sex and IQ will be correlated when conditioned on college admission because, on average, men need a higher IQ to be admitted to college. Section 3.4 further explains this example.

Why does the rescaling bias exist in logit and not in OLS? For the estimation of the logit model, the distribution of the residuals must be fixed. Generally, it's fixed as the standard logit distribution. Thus, the "true" residuals are transformed/rescaled into a standard logit distribution, and the rescaling will affect the coefficients of the model. For example, take model (4), its "true" residuals u_i^b and its "true" probabilities p_i^b. To estimate model (4), a $g()$ function, capable of transforming u_i^b into $\hat{u}_i^b = \varepsilon \sim logit\left(0, \frac{\pi^2}{3}\right)$, is applied on both sides, as illustrated in equation (5). Indeed, equation (4) is equivalent to equation (5). As a side effect, $g()$ also modifies all parameters of model (4). Thus, if a variable z_1, correlated with y, is added to equation (4), the "true" u_i^b distribution will change and so will the $g()$ function, leading to different rescaling and consequently changing the coefficients. One more nuanced must be stated: the $g()$ will only change, and OVB is only created, if $corr(\hat{\boldsymbol{\upsilon}}, x|y = 1) \neq 0$. Otherwise, the log-odds distribution will remain unchanged, resulting in no rescaling difference.

$$g\left(ln\left[\frac{p_i^b}{1-p_i^b}\right]\right) = g\left(\beta_0^b + \beta_1^b x_i + u_i^b\right) \tag{5}$$

3.1 Decomposition Procedure

Decompose the 1st source of bias: the confounding effects

$$z_i^z = \hat{\Gamma}_0^z + x_i\hat{\Gamma}_1^z + \hat{\upsilon}_i^z \tag{6}$$

To divide the confounding effects, estimate auxiliary regressions (6), where $z \in [1:p]$. Together with parameters from the full model (3), calculate $\sum_{z=1}^{p}\left[\hat{\Gamma}_1^z\hat{\beta}_2^z\right]$. This part of the decomposition is the same as Gelbach (2009). However, in antinomy with linear models, confounding effects won't account for the full bias, only the difference between the full model and the Residual Equation, model (RE) (7). The Residual Equation (first used by Karlson, Holm, and Breen (2012)) is the base equation (4) plus all the estimated residuals from the auxiliary regressions (6):

$$ln\left(\frac{P^{RE}(y=1|x,\hat{\boldsymbol{\upsilon}})}{P^{RE}(y=0|x,\hat{\boldsymbol{\upsilon}})}\right)=\hat{\beta}_0^{RE}+\hat{\beta}_1^{RE}x_i+\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\upsilon}}, \tag{7}$$

Thus, the confounding effects can explain and divide:

$$\hat{\beta}_1^{RE}-\hat{\beta}_1^{f}=\underbrace{\sum_{z=1}^{p}\left[\hat{\Gamma}_1^z\hat{\beta}_2^z\right]}_{\text{Confounding}}$$

Why is the Residual Equation free from rescaling effects? By adding the residuals of the auxiliary regressions to the base equation (4), the RE maintains the distribution of the residuals, $u_{RE}\equiv u_{full}$. The rescaling of the full model (3) and the RE (7) is the same, so there are no rescaling effects from one model to the other. Thus, equation (8) shows that we can divide the bias into the rescaling from the base model to the RE, and apply the Gelbach from the RE to the full.[8]

$$\underbrace{\hat{\beta}_1^{b}-\hat{\beta}_1^{f}}_{\text{Total bias}}=\underbrace{\hat{\beta}_1^{b}-\hat{\beta}_1^{RE}}_{\text{Rescaling}}+\underbrace{\sum_{z=1}^{p}\left[\hat{\Gamma}_1^z\hat{\beta}_2^z\right]}_{\text{Confounding}}. \tag{8}$$

Decompose the 2nd source of bias: the rescaling effects

To divide the bias between the base (4) and RE (7) models, I use the estimated residuals from the unconditional auxiliary regressions (henceforth $\hat{\boldsymbol{\upsilon}}$ or $\hat{\upsilon}^z$) to estimate conditional auxiliary regressions (CARs):

$$\hat{\upsilon_i^z}=\hat{\delta}_0^z+\hat{\delta}_1^z x_i+\hat{\delta}_2^z y_i+\hat{e}_i^z,\ \forall z\in[1:p] \tag{9}$$

Assuming that $\hat{\boldsymbol{\upsilon}}$ follows a multivariate normal distribution conditional on y and x with a precision matrix

$$\boldsymbol{\Sigma}_{p\times p}^{-1}=\begin{bmatrix}\xi_{11} & - & - & - \\ \xi_{12} & \xi_{22} & - & - \\ ... & ... & ... & - \\ \xi_{1p} & \xi_{2p} & ... & \xi_{pp}\end{bmatrix},$$

[8] A specificity of the residual equation is that the $\hat{\Lambda}$'s are equal to the respective estimated variable's coefficients in the full model, $\hat{\beta}_2$.

then the bias conditional on the state of y in $\widehat{\beta}_1$ (the bias emerging from the rescaling) is:

$$\widehat{rescaling\ bias}_{\beta_1} = \sum_{z=1}^{p} \left[\hat{\delta}_2^z \hat{\delta}_1^z \hat{\xi}_{zz} + \hat{\delta}_2^z (\hat{\Lambda}^z - \hat{\delta}_2^z \hat{\xi}_{zz}) \right], \tag{10}$$

with the constraint:

$$\Leftrightarrow \hat{\xi}_{zz} = \frac{\hat{\Lambda}^z - \sum_{\dot{z} \neq z}^{p} \hat{\delta}_2^{\dot{z}} \hat{\xi}_{z\dot{z}}}{\hat{\delta}_2^z}, \forall\, z \in \{1 : p\}; \tag{11}$$

The bias is dividable amongst **Z**. See Section A.3.9 for details on the estimation of the precision matrix of a multivariate normal distribution. The caveat is the assumption of conditional normality of $\hat{\boldsymbol{v}}$. In practice, no residual is perfectly normal, so the decomposition is an approximation that will be better or worse depending on the conditional normality of $\hat{\boldsymbol{v}}$.

Special case: independence of the residuals $\hat{\boldsymbol{v}}$

Whenever all the residuals are independent from each other, whether conditionally or unconditionally, the bias equation simplifies to:

$$\widehat{rescaling\ bias}_{\beta_1} = \sum_{z=1}^{p} \left[\hat{\delta}_1^z \hat{\Lambda}^z \right]. \tag{12}$$

3.2 A Proof for the Decomposition of the Rescaling Effects

This section provides formal proof of the rescaling decomposition. For the confounding effects, the proof was done by Gelbach (2009). Nonetheless, see Annex A.3.8 for a deduction applied to my minimum wage case.

I use L. Lee (1980) as a starting point. The decomposable difference between the RE (7) and the base equation (4) is found by using Bayes' Theorem, assuming a conditional normal distribution for $\hat{\boldsymbol{v}}$, and estimating the CARs.

Bayes' Theorem

By applying Bayes' theorem to the estimated probabilities of the RE model (7) we get:

$$\frac{P^{RE}(y=1|x,\hat{\boldsymbol{v}})}{P^{RE}(y=0|x,\hat{\boldsymbol{v}})} = \frac{P(\hat{\boldsymbol{v}}|x,y=1)}{P(\hat{\boldsymbol{v}}|x,y=0)} \frac{P^b(y=1|x)}{P^b(y=0|x)} \tag{13}$$

$$\Leftrightarrow ln\left(\frac{P^b(y=1|x)}{P^b(y=0|x)}\right) = ln\left(\frac{P^{RE}(y=1|x,\hat{\boldsymbol{v}})}{P^{RE}(y=0|x,\hat{\boldsymbol{v}})}\right) + ln\left(\frac{P(\hat{\boldsymbol{v}}|x,y=0)}{P(\hat{\boldsymbol{v}}|x,y=1)}\right). \tag{14}$$

Conditional Multivariate normality of $\hat{\boldsymbol{v}}$

If $\hat{\boldsymbol{v}}$ follows a conditional multivariate normal distribution conditional on x and y, $\hat{\boldsymbol{v}}|x, y \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, its probability density function is:

$$P(\hat{\boldsymbol{v}}|x,y) = f_{\boldsymbol{v}}(\hat{v}^1, ..., \hat{v}^p|x,y) = \frac{\exp\left(-\frac{1}{2}(\hat{\boldsymbol{v}}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{v}}-\boldsymbol{\mu})\right)}{\sqrt{(2\pi)^p det\boldsymbol{\Sigma}}}, \tag{15}$$

where $\boldsymbol{\Sigma}_{p\times p}$ is the variance-covariance matrix of CARs' residuals' $\hat{e}_i$ (making $\boldsymbol{\Sigma}^{-1}$ a positive definite and symmetric[9] precision matrix[10]); p is the number of control variables; det is the determinant; and T is the transpose indicator. The matricial form:

$$\boldsymbol{\Sigma}^{-1}_{p\times p} = \begin{bmatrix} \xi_{11} & - & - & - \\ \xi_{12} & \xi_{22} & - & - \\ ... & ... & ... & - \\ \xi_{1p} & \xi_{2p} & ... & \xi_{pp} \end{bmatrix} ; \; \hat{\boldsymbol{v}}_{p\times 1} = \begin{bmatrix} \hat{v}^1 \\ \hat{v}^2 \\ ... \\ \hat{v}^p \end{bmatrix} ; \; \boldsymbol{\mu}_{p\times 1} = \begin{bmatrix} \mu^1 \\ \mu^2 \\ ... \\ \mu^p \end{bmatrix}$$

After estimating the conditional auxiliary regressions (9) (CARs), and assuming $\hat{\boldsymbol{v}}$ is jointly normally distributed conditional on x and y[11], then we can substitute the average value by $\mu^z = \hat{\delta}^z_0 + \hat{\delta}^z_1 x + \hat{\delta}^z_2 y$ in equation (15). Furthermore. we specify the states of y in equation (15) and divide them. Therefore, we get equation (17) (for a detailed proof of this step, see Annex A.5.1):

$$\frac{P(\hat{\boldsymbol{v}}|x,y=0)}{P(\hat{\boldsymbol{v}}|x,y=1)} = e^{\frac{1}{2}\left[(\hat{\boldsymbol{v}}-\hat{\boldsymbol{\delta}}_0-\hat{\boldsymbol{\delta}}_1\mathbf{X}-\hat{\boldsymbol{\delta}}_2)^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{v}}-\hat{\boldsymbol{\delta}}_0-\hat{\boldsymbol{\delta}}_1\mathbf{X}-\hat{\boldsymbol{\delta}}_2)-(\hat{\boldsymbol{v}}-\hat{\boldsymbol{\delta}}_0-\hat{\boldsymbol{\delta}}_1\mathbf{X})^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{v}}-\hat{\boldsymbol{\delta}}_0-\hat{\boldsymbol{\delta}}_1\mathbf{X})\right]} \tag{16}$$

$$\Leftrightarrow ln\left(\frac{P(\hat{\boldsymbol{v}}|x,y=0)}{P(\hat{\boldsymbol{v}}|x,y=1)}\right) = \sum_{z=1}^{p}\left[\hat{\delta}^z_2 \sum_{\dot{z}=1}^{p}((\hat{\delta}^{\dot{z}}_0 + \hat{\delta}^{\dot{z}}_1 x + \frac{\hat{\delta}^{\dot{z}}_2}{2} - \hat{v}^{\dot{z}})\xi_{z\dot{z}})\right], \tag{17}$$

By plugging equations (17), the RE model (7), and the base model (4) into equation (14) we can see the relationship between the base equation and the RE:

$$\hat{\beta}^b_0 + \hat{\beta}^b_1 x_i = \hat{\beta}^{RE}_0 + \hat{\beta}^{RE}_1 x_i + \sum_{z=1}^{p}\hat{v}^z_i\hat{\Lambda}^z + \sum_{z=1}^{p}\left[\hat{\delta}^z_2 \sum_{\dot{z}=1}^{p}((\hat{\delta}^{\dot{z}}_0 + \hat{\delta}^{\dot{z}}_1 x + \frac{\hat{\delta}^{\dot{z}}_2}{2} - \hat{v}^{\dot{z}})\hat{\xi}_{z\dot{z}})\right] \tag{18}$$

[9] All inverse of symmetric matrices are also symmetric.

[10] If a covariance matrix has det $\Sigma \neq 0$, automatically avoiding the problem of det $\Sigma = 0 \Rightarrow \nexists\Sigma^{-1}$.

[11] If x_i is a discrete variable, this condition implies normality for all levels of x.

To obtain the expression of the bias from rescaling, we can take equation (18) and keep only the terms that interact with x:

$$x_i\widehat{\beta}_1^b = \widehat{\beta}_1^{RE}x_i + \sum_{z=1}^{p}\left[\widehat{\delta}_2^z\sum_{\dot{z}=1}^{p}\widehat{\delta}_1^{\dot{z}}x_i\widehat{\xi}_{z\dot{z}}\right] \tag{19}$$

$$\Leftrightarrow x_i\widehat{\beta}_1^b = x_i\left(\widehat{\beta}_1^{RE} + \underbrace{\sum_{z=1}^{p}\left[\widehat{\delta}_2^z\sum_{\dot{z}=1}^{p}\widehat{\delta}_1^{\dot{z}}\widehat{\xi}_{z\dot{z}}\right]}_{y-conditional\ bias}\right) \tag{20}$$

$$\Leftrightarrow \widehat{rescaling\ bias}_{\beta_1} = \sum_{z=1}^{p}\left[\widehat{\delta}_2^z\sum_{\dot{z}=1}^{p}\widehat{\delta}_1^{\dot{z}}\widehat{\xi}_{z\dot{z}}\right] \tag{21}$$

And separating $diag\left(\widehat{\Sigma}^{-1}\right)$ from the rest of the precision matrix:

$$\Leftrightarrow \widehat{rescaling\ bias}_{\beta_1} = \sum_{z=1}^{p}\left[\underbrace{\widehat{\delta}_2^z\widehat{\delta}_1^z\widehat{\xi}_{zz}}_{Individual} + \underbrace{\widehat{\delta}_2^z\sum_{\dot{z}\neq z}^{p}\widehat{\delta}_1^{\dot{z}}\widehat{\xi}_{z\dot{z}}}_{Interaction}\right] \tag{22}$$

We cannot immediately substitute the ξ's from the estimated precision matrix because we need to warrant that $\boldsymbol{\upsilon}$ does not influence the base equation (4), for logical consistency. Thus, by using equation (18) and guaranteeing that $\boldsymbol{\upsilon}$ does not influence the base equation (4) (proof in Annex A.5.2), we get this constraint:

$$\xi_{zz} = \frac{\widehat{\Lambda}^z - \sum_{\dot{z}\neq z}^{p}\widehat{\delta}_2^{\dot{z}}\widehat{\xi}_{z\dot{z}}}{\widehat{\delta}_2^z} \tag{23}$$

I suggest solving this equation by estimating the precision matrix (see Annex A.3.9 for details on the estimation of the precision matrix) and then calculating the diagonal from the constraint in equation (23). I'm attempting to find a $\widehat{\boldsymbol{\upsilon}}$ with determined levels of within correlation which ensures that $\widehat{\boldsymbol{\upsilon}}$ never influences the base equation directly. This correlation interpretation is based on the equation $\rho_{z\dot{z}} = -\frac{\xi_{z\dot{z}}}{\sqrt{\xi_{zz}\xi_{\dot{z}\dot{z}}}}$, where $\rho_{z\dot{z}}$ is Pearson's coefficient of partial correlation of the CARs residuals $corr(\widehat{e}^z, \widehat{e}^{\dot{z}})$ (Lauritzen 1996). Substituting the interaction part of equation (22) by the constraint (23), we get our final equation:

$$\widehat{rescaling\ bias}_{\beta_1} = \sum_{z=1}^{p} \left[\hat{\delta}_2^z \hat{\delta}_1^z \xi_{zz} + \hat{\delta}_2^z (\hat{\Lambda}^z - \hat{\delta}_2^z \xi_{zz})\right],$$

which is still subject to the constraint in equation (23).

$\hat{\boldsymbol{v}}$ Independent from each other

The econometrician, given specific cases, may assume that $\hat{\boldsymbol{v}}$'s are conditional uncorrelated between them. This implies that $cov(\hat{e}^z, \hat{e}^{\dot{z}}) = 0$. Also, because $corr(\hat{e}^z, \hat{e}^{\dot{z}}) = -\frac{\xi_{z\dot{z}}}{\sqrt{\xi_{zz}\xi_{\dot{z}\dot{z}}}}$:

$$cov(\hat{e}^z, \hat{e}^{\dot{z}}) = 0 \Rightarrow \xi_{z\dot{z}} = 0$$

$$\Downarrow$$

$$\begin{cases} \xi_{z\dot{z}} = 0 \Rightarrow \xi_{zz} = \frac{\hat{\theta}^z}{\hat{\beta}_2^z} = \frac{1}{\sigma_z^2} \\ \therefore \widehat{rescaling\ bias}_{\beta_1} = \sum_{z=1}^{p} \left[\hat{\delta}_1^z \hat{\Lambda}^z\right] \end{cases} \tag{24}$$

In sum, if the events are independent, the bias division is similar to Gelbach (2009) method, the only difference being that y is added to a second-wave auxiliary regression with $\hat{\boldsymbol{v}}$ as the dependent variable.

3.2.1 Final Remarks on the Decomposition

Why Estimate the Residual Equation?

It is not necessary to estimate the Residual Equation for the decomposition. Just like estimating the base equation in the Gelbach decomposition, it is only useful to confirm if the decomposition is correct. In the logit case, RE can be used to check if the division of rescaling and confounding was done correctly.

Why perform a 2-step decomposition?

If $\boldsymbol{Z}$ follows a jointly normal distribution, conditional on **X** and **Y**, applying the CAR method directly between the base model (4) and the full model (3) would be correct. However, there are several advantages in performing the logit decomposition in 2 steps, first with Gelbach (2009) decomposition and then with the CAR method. Firstly, in practice, $\boldsymbol{Z}$ will not perfectly follow a jointly normal distribution conditional on **X** and **Y**. Thus, doing the decomposition in 2 steps will make this imperfection only affect the CAR part. Secondly, since $\boldsymbol{Z}$ is unlikely to perfectly follow a jointly normal distribution, passing $\boldsymbol{Z}$ through the 1st auxiliary regressions (6) may have a *normalization* effect, i.e., $\hat{\boldsymbol{v}}$ will likely be more normal than $\boldsymbol{Z}$ (especially if **X** is continuous). The closer $\hat{\boldsymbol{v}}$/**Z** get to a normal distribution, the more accurate the decomposition. Thirdly, $\hat{\boldsymbol{v}}$ will likely have less *within-correlation* than $\boldsymbol{Z}$, because the correlation through $\mathbf{X}$ was removed, meaning that the estimation of the precision matrix is less susceptible to errors. Lastly, there are extra insights from dividing the confounding bias from the rescaling bias.

The normality caveat

The big assumption and caveat of this decomposition is the conditional multivariate distribution of $\hat{\boldsymbol{v}}$.[12] Not fulfilling this contingency will leave some parts of the rescaling bias unaccounted for. This assumption is not for statistical inference. It's to grant rigor to the CAR decomposition (the decomposition of the rescaling effects).

3.3 A Simulation

This section shows what happens to the coefficients in logit models when uncorrelated variables are added. I follow the directives of Train (2009) about simulations for binary settings. I create an artificial dataset about the effects of sex, IQ, and wealth on the probability of individuals going to college (as mentioned in Mood (2017)). Consider the following data generation process of $n = 100000$ observations:

$$college_i = \mathbf{1}[\delta_1 sex_i + \delta_2 iq_i + \delta_3 wealth_i + \varepsilon_{it} > 0], \tag{25}$$

where sex_i is 1 for men and 0 for women, $iq_i \wedge wealth_i \sim N(0,4)$ and $\varepsilon_i \sim logit\left(0, \frac{\pi^2}{3}\right)$, a standard logit distribution. I use the Cholesky decomposition to set the covariances. Table 1 shows the results of estimating the simulated data with a logit model trough 3 cases:

- Case 1: sex_i is correlated with iq_i and $wealth_i$. All the covariates are positive and $\delta_1 \wedge \delta_2 \wedge \delta_3 = 1$.
- Case 2: sex is uncorrelated with the other covariates, but correlated conditional on y=1. $cov(sex, iq) = 0$ and $cov(sex, wealth) = 0$ and $\delta_1 \wedge \delta_2 \wedge \delta_3 = 1$. This is the most realistic situation in most of the developed world.
- Case 3: sex is uncorrelated with the other covariates, both unconditional and conditional on $y = 1$. There are several ways to achieve this setting $cov(sex, iq) = 0$ and $cov(sex, wealth) = 0$ and $\delta_1 \wedge \delta_2 \wedge \delta_3 = 1$. The easiest way to achieve this is by setting the covariances to 0, $\delta_1 = 1$, $\delta_2 = 0$, and $\delta_3 = 0$.

Table 1 also includes "non-normality" rows. In these, iq_i and $wealth_i$ are turned into binary variables, with 0 as the threshold.

[12] Because the normal distribution is a stable distribution, linear combinations of it are a $P(\upsilon^1, \ldots, \upsilon^p)$ jointly normal distribution (Abdul-Hamid and Nolan 1998). Therefore, to be more precise, the base assumption is that each individual $\hat{\upsilon}^z$ is a normal distribution.

Table 1: Results from the Simulation

	sex_i						
	Estimated Coefficients			Estimated Biases		Decomposable Biases	
	Base	RE	Full	Confounding	Rescaling	Confounding	Rescaling
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Case 1							
With normality							
Logit	1.24	2.40	1.01	1.39	-1.16	1.39	-1.13
LPM	0.15	0.15	0.06	0.09	0.00	0.09	0.00
Without normality							
Logit	0.68	0.75	1.00	-0.25	-0.08	-0.25	-0.03
LPM	0.11	0.11	0.15	-0.04	0.00	-0.04	0.00
Panel B: Case 2							
With normality							
Logit	0.36	1.00	1.00	0.00	-0.64	0.00	-0.62
LPM	0.08	0.08	0.08	0.00	0.00	0.00	0.00
Without normality							
Logit	0.87	1.00	1.00	0.00	-0.12	0.00	-0.05
LPM	0.13	0.13	0.13	0.00	0.00	0.00	0.00
Panel C: Case 3							
With normality							
Logit	1.00	1.00	1.00	0.00	0.00	0.00	0.00
LPM	0.14	0.14	0.14	0.00	0.00	0.00	0.00
Without normality							
Logit	0.99	0.99	0.99	0.00	0.00	0.00	0.00
LPM	0.14	0.14	0.14	0.00	0.00	0.00	0.00

Notes: This table shows the result of estimating the data generation process in equation 25 with a logit and linear probability model. In Panel A, all variables are correlated. In Panel B, variables can only be correlated conditional on college attendance. In Panel C, the variables are uncorrelated, both unconditional and conditionally on the state of the dependent variable. The tab "Estimated Coefficients" shows the coefficient of sex_i in a regression without any covariates (column 1); with the uncorrelated residuals from iq_i and $wealth_i$ (column 2), and with iq_i and $wealth_i$ (column 3). Columns (4) and (5) are the biases. (4) = (2) - (3) and (5) = (1) - (2). Column (6) is the bias explained by the Gelbach decomposition. It's the sum of the components of $wealth_i$ and iq_i. Column (7) is the part explained by the CAR decomposition, which is also the sum of the $wealth_i$ and iq_i components of the bias.

Table 1 starts with "Panel A: Case 1". Column (4) shows that confounding effects exist both in logit models and in linear probability models (LPM). But the rescaling effects, the bias component from removing uncorrelated variables, only exists in the logit model. The sum from the Gelbach decomposition is reported in column (6). It manages to account for all confounding effects in both models. On the rescaling bias, while decomposable rescaling is decomposable using the CAR method, it's only precise if the variables are normally distributed (because that makes $\widehat{v}$'s very likely to be normally distributed conditional on sex_i and $college_i$). Notice that, percentage wise, without normality, the CAR decomposition accounts for less than 35% of the rescaling bias, while with normality almost the full bias is accounted for.

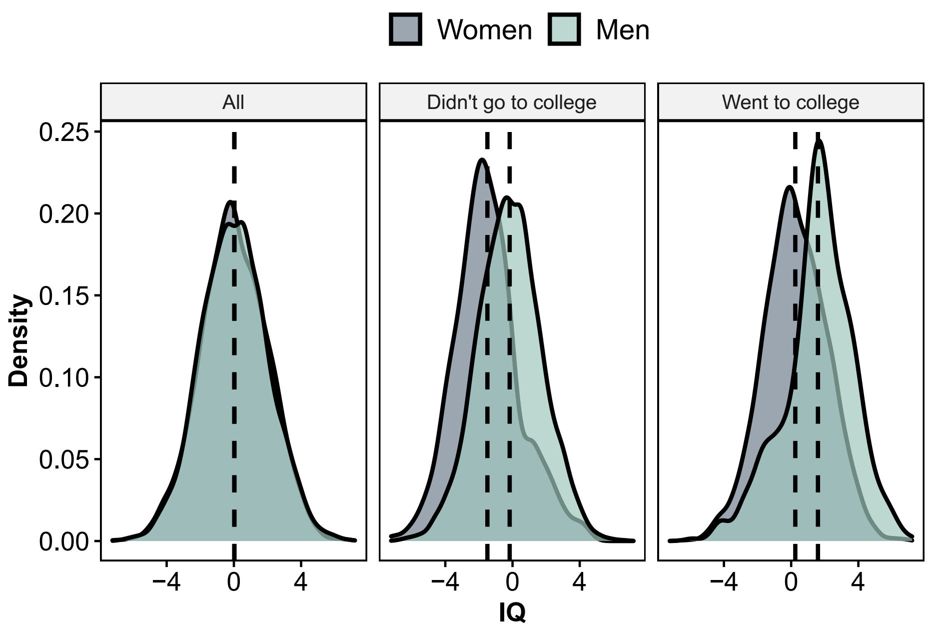
Table 1 “Panel B: Case 2” shows that when the variables are uncorrelated, there is no bias in the linear model but there is bias in the logit model. The only existing bias is from the logit rescaling; and, again, only if the normality assumption renders rigor to the CAR decomposition.

Finally, Table 1 “Panel C: Case 3” shows a case where no model is biased. By setting the correlation to 0, both conditional and unconditional on the state of y, both models are unbiased.

3.4 Interpretation of the Logit Decomposition

The decomposition of the confounding and the rescaling effects have different interpretations. Decomposing confounding effects has the same interpretation as the Gelbach (2009) decomposition in linear models. By accounting for how much each variable from $\mathbf{Z}$ explains the sensitivity of $\mathbf{X}$, we can conclude how much of $\mathbf{X}$'s average effects ($\hat{\beta}_1^{\,f}$) are explained by each $\mathbf{Z}$, individually.

Figure 1: Distribution of IQ Conditional on Going to College, by Gender



Notes: This plot shows the distribution of the iq_i variable from the second case of the data generation process in equation 25, by sex. The panel on the left shows the unconditional distributions. The central panel shows the distributions conditional on $y = 0$, i.e., not going to college. The right-hand panel shows the distribution conditional on $y = 1$, i.e., not going to college.

On the other hand, decomposing rescaling effects involves dividing the impact controls have on the heterogeneity of the treatment effect on the treated. By comprehending which variables create the biggest variability in the treatment effect, we can understand the main determinants of that treatment.[13]

The rescaling effects result from a correlation between $\hat{\boldsymbol{\upsilon}}$ and x, conditional on the state of y being 1. Because $\hat{\boldsymbol{\upsilon}}$ and x are residuals of the auxiliary regressions (6), they would be uncorrelated if they were not conditioned on the state of y. Therefore, the CAR decomposition already measures correlation that occurs *only* when conditional on the state of y.

The CAR decomposition is conditioning the correlation of x and $\hat{\boldsymbol{\upsilon}}$ on the state of y, therefore conditioning on the state of the outcome. Although the effect of treatment on the *log-odds* is the same, the *log-odds* have a non-linear impact on the probability of y in logit, concretely $\frac{exp(\text{log-odds})}{1+exp(\text{log-odds})}$. Thus, by observing the correlation following a threshold screening ($y = 1$), we can assess the heterogeneity of the *log-odds* effects actually manifesting in the outcome variable. Some variables will be more decisive than others. Thus, if a variable is extremely important for treatment to materialize as a change in y, there should be a high conditional correlation. Otherwise, if a variable is irrelevant for the heterogeneity, the correlation should be 0 after conditioning on the state of y, because treatment affected all observations the same way and the pattern was not changed. This characterizes heterogeneity and allows us to answer the following question: How much does each variable matter to explain the heterogeneous impact of **X** on **y**? (However, we do not detect heterogeneity of treatment in the *log-odds* nor heterogeneity of the *log-odds* effect in the predicted probabilities. Instead, we capture what variables matter to make treatment exacerbated effects on the actual outcome.)

To clarify, I plot simulated data from the data generation process in equation (25). I use Case 2 and focus on the relationship between IQ and sex. In Case 2, we make the (realistic) assumption that IQ and sex are uncorrelated and both that matter to predict the probability of an individual going to college. Figure 1 shows density plots of the IQ variable by sex. The distribution and average IQ are identical for men and women, as would be expected. Nonetheless, conditional on going to college, men have a higher average IQ than women. Same for individuals who didn't go to college. This peculiar situation emerges from the fact that, because men are less prone to go to college, they need other factors to pass the threshold of variables that determined college admission. Thus, there will be a correlation conditional on the state of y.

How does this correlation conditional on the state of y (going to college, in the example) embody heterogeneity? Because being a man (treatment) has an effect dependent on IQ and wealth. To some men, because their IQ is lower, being a man will demote them from going to college. Others, because their IQ is high, won't be demoted. The same effect of the *log-odds* is affecting men differently in

[13] Sometimes, this relationship is so important for the fit of the model that researchers add interaction terms between x and some variables of **Z**. However, interactions will tell us how the effect of x on y changes conditional on **Z**, not how important each **Z** is at explaining the heterogeneity of the treatment effect.

the translation of *log-odds* to the probabilities. In other words, the causal effect of being a man is the same in both (note that we are not seeing the effect on the *log-odds*), but the interaction with other variables will determine if sex was a decisive factor. Is IQ or wealth a more important factor making sex a decisive factor? That is the heterogeneity being measured. In the next section, I show an application of the logit decomposition to the minimum wage case.

4 An Application: the Heterogeneity of the Minimum Wage Effect on Displacement

What are the determinants of the displacement effects of the minimum wage (MW)? The heterogeneous response to MW hike can be divided with the CAR logit decomposition. For that, I analyse a natural experiment in Portugal.

From 1974 to 1986, 18- and 19-year-old workers had a lower MW. It was 75% of the MW binding to $20+$ year-old workers. In 1987, that lower MW was eliminated for constitutional reasons. Thus, in 1987, only some MW workers had a MW increase. I find that the MW increased the probability of displacement by 6 percentage points. I further find, by using the CAR logit decomposition, that the heterogeneity in the MW impact on separations is 65% explained by firms, 28% by the worker, and 7% by tenure. Thus, the firm where a worker is employed is the most important factor in determining whether or not he will be displaced because of the MW.

The identification strategy of the MW effect on separations consists in building a logit model employing a Triple difference methodology (DDD) to identify the MW effect. The DDD compares workers on three dimensions. (1) Wage: MW workers vs. above MW workers; (2) Age: Teen workers vs. Young workers; (3) Time: at treatment and post-treatment. This strategy isolates shocks that are specific to teen MW workers. The controls include tenure, firm size, firm and worker high-dimensional fixed effects (FE).

Portuguese Minimum Wage Literature

In Portugal, three papers have analysed the 1987 MW natural experiment with a Difference in Differences methodology, with the following main results:[14]

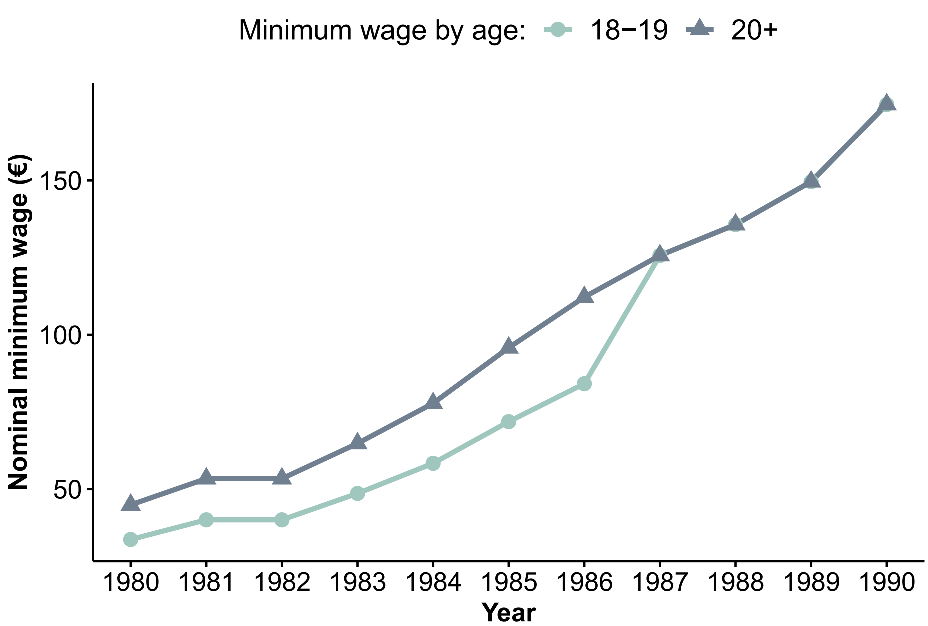
- Pereira (2003) finds a negative employment-MW elasticity between -0.2 and -0.4 for teenagers and a positive employment spillover for 20- to 24-year-old workers. I have, like Portugal and Cardoso (2006), tried to replicate Pereira (2003) and we couldn't reach her results. I reached negative elasticities, but closer to zero;

[14] In Annex A.4 I further discuss Portugal's specific wage-setting mechanism that can influence the MW effects.

- Portugal and Cardoso (2006) find null employment effects for teenagers because both hiring and separations decreased. Portugal and Cardoso (2006) is particularly important for my thesis because they also use a logit model on separations with forward years as counterfactuals;[15]
- Cerejeira (2008) finds negative employment elasticities between -0.42 and -0.47 for the directly affected teens, and substitution of low-skill teens for teens above the MW. His results manage to reconcile a negative effect on the separations of teen MW workers and the null overall effect found by Portugal and Cardoso (2006).

The distinctive points of this thesis identification strategy are twofold. Firstly, I build a Triple Difference (DDD) estimator, instead of a Difference in Differences. This allows me to account for teen-specific trends. In Annex A.3.1, I show their importance to this analysis. Isolating teen MW workers avoids misclassification issues by not binding together workers that were treated and workers that weren't (Jardim et al. (2022); Dube (2018)). Secondly, I have the goal of detecting the main determinants of the heterogeneous impacts of the minimum wage on separations, within affected workers (within the teen MW workers of 1986). See Annex A.6 for a deeper literature review, of where this study fits in the literature.

Figure 2: Evolution of the Minimum Wage in Portugal



Notes: This figure shows the evolution of the nominal national minimum wage of Portugal, from 1980 until 1990, and for all workers at the age of majority. The minimum wage applicable to 18- to 19-year-old workers was 75% of the MW applicable to 20+ year-old workers. We can see that the difference ended in 1987. After 1990, 18- and 19-year-old workers have continued having the same minimum wage as non-teen workers. Only applicable to continental Portugal since the autonomous regions of Madeira and Azores have an independent minimum wage policy. Source: Instituto Nacional de Estatística (INE).

[15] Portugal and Cardoso (2006) compare separations in the treatment years with separations after and before. In section A.2.2 I explain why I don't use years before 1986.

4.1 The Natural Experiment

In 1974, the MW law was introduced in Portugal, covering all workers 20 years old or more. Since then, the MW has undergone several nominal increases (Figure 2) and changes to its age coverage. Table 2 reports the applicable national MW by age, compared to the adult MW. This thesis focuses on the bold area of Table 2: when 18- and 19-year-old workers went from receiving 75% of the MW in 1986 to receiving its full amount in January 1987.[16] This policy shift is econometrically relevant, considering the following aspects:

- **It's a natural experiment** that creates a wedge by having different MW increases *within* MW workers. MW workers at 20 years old and above became a natural counterfactual for teen MW workers;
- **It's an exogenous shock** imposed by the Constitutional Court;
- **It's a big increase**, illustrated in Figure 2. In real terms, the growth rate of the MW was 36.25% for teens (and only 2.19% for adults). The main advantage of a big increase is that it dilutes identification errors of the MW effect (see Harasztosi and Lindner (2019) analysis of a 60% MW increase in Hungary). Figure 3 shows the MW impact on the Kaitz index of teens (Kaitz 1970). After the MW increase, the Kaitz index escalated for teens despite decreasing for adults.
- **There is a small overall incidence**: only 0.31% of the workforce was affected. Accordingly, unemployment effects coming from a decrease in production are less plausible (Gregory and Zierahn 2020). Therefore, while workers waging close to the minimum wage may suffer a substitution effect (Cengiz et al. 2019), the upper end of the distribution should remain uncontaminated, making it a good counterfactual to measure teen trends.

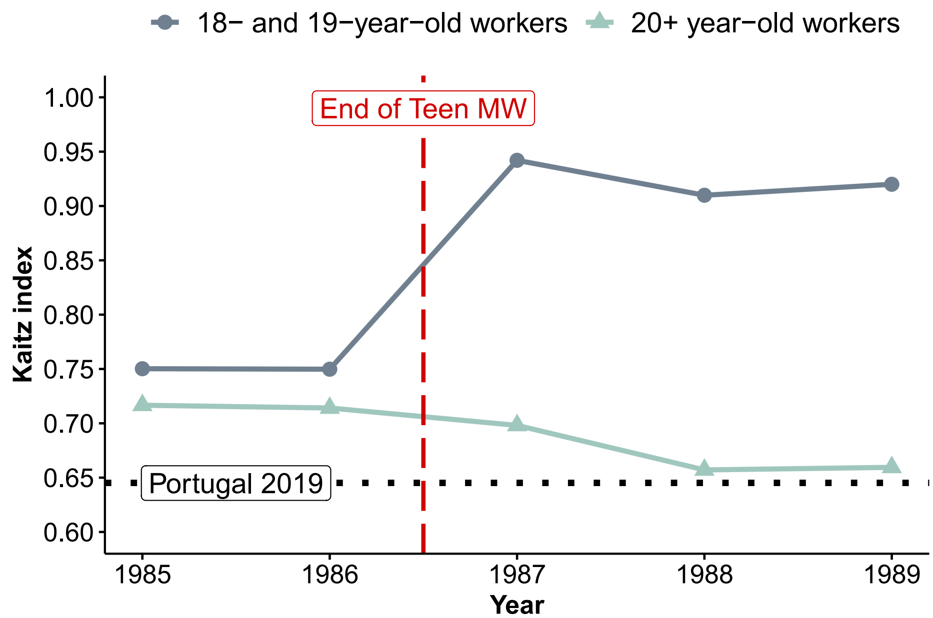
Table 2: Coverage of the Minimum Wage by Age

Year \ Age	15	16	17	18	19	20+
1979-1986	50%	50%	50%	**75%**	**75%**	**100%**
1987	50%	50%	75%	**100%**	**100%**	**100%**
1988-1997	75%	75%	75%	100%	100%	100%
1998-now	100%	100%	100%	100%	100%	100%

Notes: This table shows the share of the minimum wage received by workers according to their age and year. For example, workers that were 15 years old in 1990 were entitled to 75% of the national minimum wage. In bold are the time frame and age bins of the natural experiment explored in this thesis. Sources: Diário da República, several issues: (1) Decreto-Lei n.º 440/79, 2^{nd} article, for the setting of the age percentages in 1979; (2) Decreto-Lei n.º 69-A/87, 4^{th} article, for the 1987 changes; (3) Decreto-Lei n.º 411/87, 4^{th} article, for the 1988 updates; (4) Lei n.º 45/98, unique article, for the elimination of all age exceptions. Exceptions to the law are detailed in Table 11 of Annex A.4.1.

[16] This legislative change had several exceptions, outlined in table 11 in Annex A.4.1.

Figure 3: Kaitz Index by Age Group in Portugal

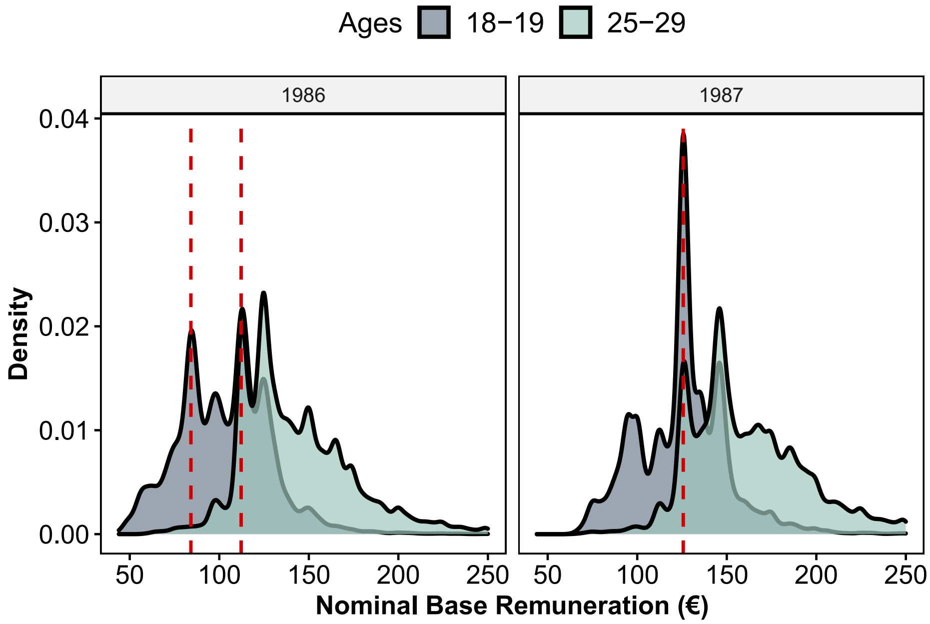


Notes: This figure shows the Kaitz index by age group, in the Portuguese private sector, from 1985 to 1989. The Kaitz index is computed as $minimum\ wage_{age}/median\ wage_{age}$. The median wage for each age group includes all compensations for extra hours and other subsidies. The sample excludes workers from Madeira or Azores, in the primary sector, with a part-time job and independently employed. The red dashed line represents the elimination of the teen MW in January 1987 (a status previously granting only 75% of the national MW to workers either 18 or 19 years old). For reference, the dotted black line is the Portuguese Kaitz index in 2019, for all age groups. Sources: Quadros de Pessoal for wage data; Instituto Nacional de Estatística, for the minimum wage data.

Figure 4 shows the distribution of nominal base wages. In 1986, Figure 4 has two dashed lines, the teen MW and MW, respectively from the left to the right-hand-side. In 1987, it has one dashed line, the unique MW to all workers at the age of majority. The increase in wages to the new MW line in 1987 is visible.[17] Furthermore, Figure 4 also shows the existence of wage spillovers: jobs paying above the MW in 1986 were also pushed to higher wages.

[17] Yet, some questions about bindingness are legitimate, given the large lumps in the 1987 panel to the left of the MW line. In Annex A.4.1, I explain the exceptions to the law and show that the MW was in fact binding.

Figure 4: The Effect of the Minimum Wage on the Wage Distribution



Notes: This figure shows the impact of the minimum wage (MW) on the wage distribution of two groups: teens (18 or 19 years old) and young workers from 25 to 29 years old. Base wages, which exclude all benefits, overtime payments, and indemnifications, are the ones bound by the MW. The sample excludes workers from Madeira or Azores, in the primary sector, with a part-time job and independently employed. The red dashed lines are the minimum wages; the left line in 1986 is the teen MW. Sources: Quadros de Pessoal for wage data; Instituto Nacional de Estatística, for the minimum wage data.

4.2 Data

Quadros de Pessoal (QP) is a linked-employer-employee-dataset annually collected by the Ministry of Employment, Solidarity and Social Security (MTSSS), with a census to all private firms with at least one employee in Portugal. QP assembles information at the establishment, firm, and worker level, with a fictitious ID for each of them (except for personnel on a long leave). Its legally mandatory nature ensures high response rates. Additionally, the Ministry's inspectors and the obligation to post the map of wages of the establishment in a public space of that establishment ensures adherence to the MW and collective agreements and the reliability of the information. Nowadays, QP collects information about more than 300 thousand firms and almost 3 million workers. See Annex A.2 for a detailed description of QP variables and Table 3 for relevant descriptive statistics for this experiment. Until 1993, QP was annually recorded in March.

The QP dataset allows a worker to be reported under two or more different firms simultaneously if an individual has several jobs. I build an algorithm to select what observations to keep, correctly dealing with separations. The details are in Annex A.2.1.

Table 3: Descriptive Statistics

Bins			Separations		ln(firm size)		Tenure		Deleted Obs. for the logit TWFE (%)				Worker Obs.	
Year	Wage	Age	mean	sd	mean	sd	mean	sd	Total	Singleton	Perfectly Classified	Separation Singleton	mean	sd
1986	MW	18-19	0.27	0.45	3.66	1.48	27.20	18.44	15.21	9.87	5.93	24.23	15.24	9.37
1986	MW	25-29	0.25	0.43	4.41	1.74	80.28	60.51	12.71	8.27	4.84	22.88	14.08	8.82
1986	Above	18-19	0.25	0.43	4.45	1.71	35.80	41.28	18.78	11.26	8.47	27.89	14.07	9.62
1986	Above	25-29	0.22	0.41	5.60	2.35	73.91	54.34	9.90	6.85	3.26	23.34	15.97	9.79
1988	MW	18-19	0.36	0.48	4.24	1.54	26.49	20.25	15.51	10.29	5.82	23.05	14.48	9.05
1988	MW	25-29	0.39	0.49	3.84	1.66	53.45	52.44	11.12	6.64	4.80	14.79	14.68	8.94
1988	Above	18-19	0.38	0.49	4.47	1.73	30.48	42.18	16.57	10.65	6.63	23.02	14.23	9.39
1988	Above	25-29	0.32	0.47	5.09	2.23	62.47	50.82	6.92	4.46	2.58	12.91	16.87	9.72

Notes: This table shows descriptive statistics relevant to the Triple Difference estimator employed in this thesis. The column ”bins” describes the characteristics of the workers considered. The sample excludes workers from Madeira or Azores, in the primary sector, with a part-time job and independently employed. A $separation = 1$ if a worker employed at firm J is no longer employed at that firm in the following year. Firm size is in natural logs and tenure is in months. The section ”Deleted Obs. for the logit TWFE” shows the percentage of observations deleted by applying the necessary restrictions for the estimation of worker and firm fixed effects. ”Singleton” refers to the percentage that is deleted because a given worker or firm only has one observation. ”Perfectly classified” refers to the percentage that is deleted because a given worker or firm only has separations or never has separations. The ”Perfectly classified” calculations already exclude singletons. Total is the sum of perfectly classified and singletons. ”Separation singletons” refers to the percentage of separations that are deleted because they are singletons (singleton elimination conditional on separation). ”Worker observations” is the number of obsevations by worker, from 1986 to 2019, by bin. Source: Quadros de Pessoal.

4.3 Identification of the Minimum Wage Effect

Equation (26) is the main specification equation. Separation years refer to the earliest year involved; for example, 1986 refers to the separations from 86 to 87. It compares workers over three dimensions: age, wage, and time; it's a triple difference methodology (DDD). A DDD is the difference between two Differences in Differences (see Olden and Møen (2022) benchmark deduction). It also contains two high-dimensional fixed effects and year fixed effects. Thus, it's an adaptation of Abowd, Kramarz, and Margolis (1999) AKM model to the binary setting. The parameter holding the main DDD result is α_{88}, interpretable as the increase in the propensity of separation which was due to the MW.

$$\begin{aligned} ln\left(\frac{Pr(y_{it}=1|\boldsymbol{X})}{1-Pr(y_{it}=1|\boldsymbol{X})}\right) = \sum_{t=1987}^{2019} & [\beta_{1t}t + \beta_{2t}teen_{it} + \beta_{3t}mw_{it} + \\ & \beta_{4t}(t \times teen_{it}) + \beta_{5t}(t \times mw_{it}) + \beta_{6t}(teen_{it} \times mw_{it}) + \\ & \alpha_t(t \times teen_{it} \times mw_{it})] + \\ & \xi_1 ln(firm\ size_{J(it)t}) + \xi_2 tenure_{it} + \theta_i + \psi_{J(it)} + \varepsilon_{it} \end{aligned} \tag{26}$$

$Where$:

$$
\begin{aligned}
t&: \text{ Discrete variable of years from 1986 to 2019}\\
i&: \text{ Worker ID}\\
J&: \text{ Firm ID}\\
teen_{it}&: \text{ Dummy variable, 1 if 18-19 years old, 0 if 25-29}\\
mw_{it}&: \text{ Dummy variable, 1 if a worker is on the MW, 0 if on the 40 percentile above MW}\\
firm\ size_{J(it)t}&: \text{ Number of employees in the firm}\\
tenure_{it}&: \text{ Number of months working in the } J \text{ firm}\\
\theta_i&: \text{ Worker fixed effects}\\
\psi_{J(it)}&: \text{ Firm fixed effects}\\
\boldsymbol{X}&: \text{ Matrix of all explanatory variables}\\
y_{it} & \begin{cases} 1 & \text{, if the worker separated in } t+1 \\ 0 & \text{, if the worker didn't separate in } t+1 \end{cases}
\end{aligned}
$$

See Annex 12.3 for a detailed description of the variables.

I must note that the basic non-interaction terms β_2 and β_3 are not interpretable. Worker and year fixed effects wipe out most of the observations.[18]

I estimate model (26) with Stammann (2018) procedure for high-dimensional two-way fixed effects (TWFE) logit models, controlling for tenure, firm size, firm FE $\psi_{J(it)}$ and, a novelty in the MW literature, worker FE θ_i; an adaptation of Abowd, Kramarz, and Margolis (1999) to separations. The estimates are also corrected to the incidental parameter problem (IPP) using the analytical bias correction derived by Ivan Fernández-Val and Weidner (2016). And because separations represent around 30% of the observations (see Table 3) I use King and Zeng (2001) rare event bias correction.

The estimation, the bias correction procedures, the computation of the average partial effects, and other details are found in Annex A.3. In the rest of this section, I justify my choice of counterfactuals for teen minimum-wage workers. The counterfactual covers 3 dimensions: *time*, *wage*, and *age*.

Time

A separation occurs when a worker i is employed at firm J in year t but isn't in the same firm in $t+1$. The binary variable y is 1 in case of separation and 0 otherwise.

The new MW law was enacted in January 1987. The data was annually collected in March.[19] Thus, fluxes from 1986 to 1987 reflect the immediate MW response, from 0 to 3 months. Fluxes from 1987 to 1988 reflect the impact from 3 to 14 months. I consider the separations from 1988

[18] See Roth et al. (2022) for the theory behind fixed effects and Differences in Differences. Furthermore, see Bossler and Gerner (2020) for an application of those principles to the minimum wage case.

[19] After 1993 it's collected in October.

to 1989 free from any treatment effect. Because α_{88} is already the difference between separations of 86/87 and 88/89, holds the MW causal effect from short-run effects. In Annex A.3.2, I test the usage of year 1991 as a counterfactual and find no statistically significant change.

I employ the DDD with forwarding years, not years before the treatment took place, because the data from 1985 has an issue that disallows its usage. See Section A.2.2 for an explanation of the data problem. I look forward to infer the past. This approach is only valid if the dependent variable (separations) is only affected by the MW in a transitory manner. Hence the usage of the logit model. Separation rates are not expected to suffer a permanent *ceteris paribus* impact from the MW (unlike the share of low-wage workers, for example), regardless of market competitiveness (see Stigler (1946) model) or the internal organization of a firm. Portugal and Cardoso (2006) have done something similar in this experiment, where, with a logit model, they compare separations of 1986 with the separations of 1988. Which implies a 1st assumption.

Assumption 1: Only transitory effects
Treatment only has a transitory effect after its implementation.

Assumption 1 substitutes the usual assumption in DDD, "no anticipatory effects", i.e., the treatment has no causal effects prior to its implementation. Here, the hypothesis that may break *Assumption 1* is that teens may become less separation prone after the MW raise, caused by either a simple lay-off of the more separation-prone (demand side) or a substitution effect (supply side) where more productive teens are encouraged to leave school as a result of the MW hike, driving out formerly employed teens (these effects were previously found by Neumark and Wascher (1995b) and Neumark and Wascher (1995a)). I include worker FE to account for this possible issue, as it allows me to account for the behavior of workers for the rest of their professional lives in the private sector.

Age

Young MW workers act as a natural counterfactual. Pereira (2003) finds an *age spillover* to 20-24 year-old workers because firms try to maintain their tenure hierarchy (Doeringer and Piore 1971), and treated teens get older and in the following years. Thus, I use workers between 25-29 years old as counterfactual. Moreover, Annex A.3.2 tests the 20-24 age bin and finds that there is no statistically significant difference.[20]

Wage

The *MW* bin comprises workers whose base wage is within a 5€ bandwidth of the MW. The

[20] A note on the age of treatment: if I was comparing with years before 1986, 19-year-old workers couldn't be used because they would have had a constant MW hike by turning 20 in the following years. However, in the time frame *after 86*, which I use, these workers suffer a unique big MW hike.

above bin includes workers earning above the 40th percentile of base wages from the distribution of workers above the MW. Furthermore, the percentiles are calculated by age bin.

This counterfactual achieves it because: (1) it's within age bins, making it time consistent, which is relevant because the MW starts at different levels by age group; (2) it's relative to the MW, reflecting the corollary of Internal Labour Markets (Doeringer and Piore 1971) which states that workers above the MW are raised for firms to maintain their internal hierarchies; (3) it's a percentile of the wages above the MW, not a percentile of the wages, reflecting the wage compression predicted by Neoclassical theory (Stigler 1946); (4) the low incidence of the MW hike leaves the upper end of the wage distribution unaffected (as discussed in Section 4.1); (5) finally, I choose the 40th threshold because teen workers earning the normal MW in 1986 (Figure 4 shows a lump there) were in the 40th percentile, and keep being near the 40th percentile in 1987, after the MW increase.[21]

The goal of the wage counterfactual is to retrieve a teen trend unaffected by the MW mechanics. Figure 10, in Annex A.3.1, shows that teen trends in separations are quite relevant.

Assumption 2: Parallel Trends
In the absence of differences in the MW growth, teens and young MW workers would have moved in parallel lines, after adjusting for teen trends. I test this in Section A.3.2.

Assumption 3: Clustered Standard Errors
I cluster the standard errors (SE) at the worker and year levels, because the treatment assignment mechanism is clustered on teen MW workers in 1987. See Abadie et al. (2017) for an overview on clustering SE and Roth et al. (2022) for the difference in differences case.[22]

4.4 Results on the Average Effect of the Minimum Wage

Table 4 shows the results for the parameter that identifies the MW effect, α_{88}. The Table reports estimated from 3 models, the full model (26), the base model (27), and the base updated model, which is the base model (27) with the same sample as the full model (26), which removes the perfectly classified and singletons. The full model (26) also has average partial effects shown in column (4). The rest of the coefficients can be seen in Annex A.3.7.

All models show an increase in separation propensity due to the MW. The average partial effect of the MW is an increase of 6 percentage points in the probability of separation. Regarding the

[21] I test if paying the full MW was a firm-level policy. I conclude that most medium/large firms held teen workers earning both MWs.

[22] For the base models, I use the two-way clustered sandwich estimator in the *R* package *sandwich*, and to the full mode with the D variable, I calculate it with Fisher Randomized tests (Fisher 1935; Davison and Hinkley 1997) and compute the confidence intervals with Hahn (1995) method.

controls, in Table 4 we can see that firm size and tenure are statistically significant and reduce the probability of separation.

Table 4: Main Estimates of the Logit models with Separations

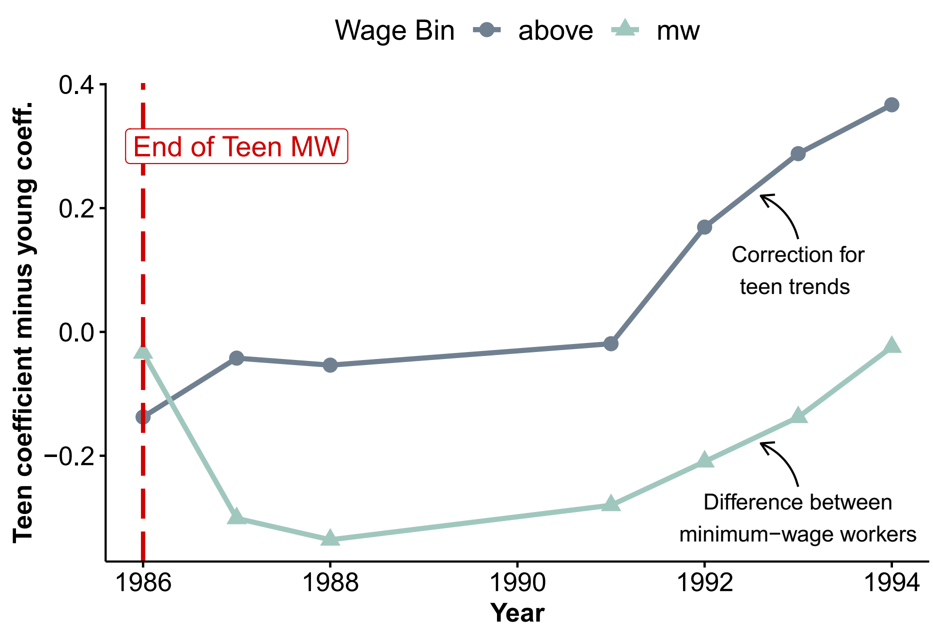
Dependent Variable:	Logit Models			APE from
Separations	Base	Base updated	Full model	Full Model
$mw \times teen \times 1988$	**0.21***	**0.36***	**0.39***	**0.06***
	$[0.11; 0.30]$	$[0.24; 0.47]$	$[0.25; 0.55]$	$[0.03; 0.08]$
$log(firm\ size)$	-	-	-0.02^{*}	-0.004^{*}
			$[-0.02; -0.02]$	$[-0.004; -0.004]$
$tenure$	-	-	-0.00^{*}	-0.0004^{*}
			$[-0.00; -0.00]$	$[-0.0004; -0.0004]$
(...)				
Worker FE	No	No	Yes	Yes
Firm FE	No	No	Yes	Yes
Updated Sample	No	Yes	Yes	Yes
Deviance	1384131.15	1152272.74	61593462.09	-
Num. obs.	1062140	59260884	59260884	-
Num. worker:	-	-	579236	-
Num. groups: firm	-	-	678054	-

Notes: * Null hypothesis value outside the confidence interval of 95%, with clustered standard errors by worker and year. This table shows coefficients from the full model 26, the base model 27, and an updated version of base model that further excludes singletons and perfectly classified observations (like the full model). Details are in Section A.2. The full model includes all the bias corrections of Section A.3.5. APE are the average partial effects of the full model, computed as shown in Annex A.3. The sample excludes workers from Madeira or Azores, in the primary sector, with a part-time job and independently employed. Source: Quadros de Pessoal.

$$
\begin{aligned}
ln\left(\frac{Pr(y=1|X)}{1-Pr(y=1|X)}\right) = \sum_{t=1987}^{2019} & [\beta_{1t} t + \beta_{2t} teen_{it} + \beta_{3t} mw_{it} \\
& + \beta_{4t}(t \times teen_{it}) + \beta_{5t}(t \times mw_{it}) + \beta_{6t}(teen_{it} \times mw_{it}) \\
& + \alpha_t(t \times teen_{it} \times mw_{it})] + \varepsilon_{it}
\end{aligned} \tag{27}
$$

Figure 5 reports the difference $teen - young$ of separation propensity within each wage bin, using estimates from the bias-corrected equation (26). Thus, a positive number implies a higher coefficient for teens than for young workers, meaning a higher separation propensity for teens. The "mw" line incorporates the treated group (teen MW workers from 1986 to 1987) and the line with workers above the MW is meant to be a correction for teen-specific trends, showing what would be the evolution of the "mw" line in the absence of the MW increase. Separations from 1986 to 1987 (identified with the red line) show that the difference was higher for MW workers than for above MW workers, indicating an increase in separations of teen MW workers due to the MW. This effect fades away after the MW shock, with the above MW workers having a consistently higher difference.

Figure 5: The Minimum Wage Effect on Separation Propensity

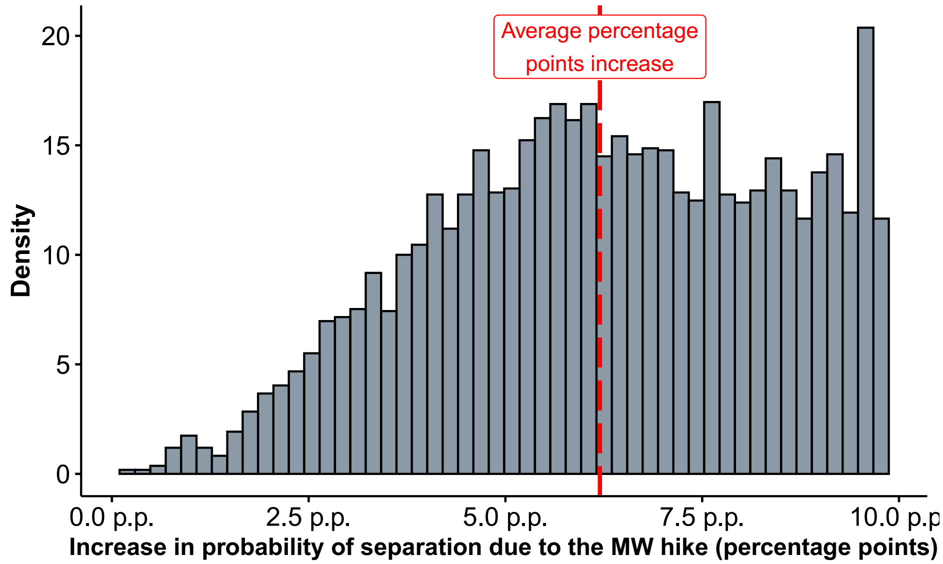


Notes: This figure reports estimates from the full model 26. The differences between the lines are the α estimates. Every dot is computed by: the logit coefficient of teens minus the logit coefficient of young workers, within each wage bin. Thus, a negative value indicates that young workers of that wage bin are more likely to separate than teen workers. The wage bins refer to workers at the minimum wage (MW) and above the 40th percentile of wages of workers above the MW. In 1986, the MW for teen workers was 75% lower, so the separations in the red line, which are separations from 86 to 87, show a MW shock to young workers. Control variables: tenure, firm size, worker and firm fixed effects. The bias corrections are found in Section Section A.3. The MW workers are defined from a bandwidth of 5€ for each side of the nominal MW. The sample excludes workers from Madeira or Azores, in the primary sector, with a part-time job and independently employed. Source: Quadros de Pessoal.

4.5 The Logit Decomposition Applied

Although I find an increase of 6 percentage points in the probability of separation because of the MW, that number, both in the logit model and reality, does not apply equally to all teen MW workers. Some workers are going to be more impacted than others. The effect might vary by firm, work ethics, *et cetera*. In Figure 6 we can visualize this heterogeneity in the logit model. Figure 6 represents the estimated impact of the MW on the treatment group. It's visible that separations increased because of the MW. Furthermore, it's clear that the non-linearity of the logit model allows differently affected workers to have a different MW impact. I use the CAR to see what are the main variables behind the different materialization of these *log-odds* effects on actual separations. In other words, what are the main determinants of whether a worker is greatly affected or just slightly affected by the MW?

Figure 6: Minimum Wage Heterogeneous Effect on Separations



Notes: This figure shows the impact of the 1987 elimination of the teen MW on the probability of separation of 18-19 MW workers. The teen MW was status previously granting only 75% of the national MW to workers either 18 or 19 years old. Its ending emanated a 36.25% increase in MW to teens at the age of majority. The probabilities are computed from the main logit specification, equation 26. The MW estimate is achieved using a Triple Difference methodology with time (the ending of teen MW and after), age (teen and young workers), and wage (MW and above MW workers). The controls include tenure, firm size, worker and firm high-dimentional fixed effects. Estimates have two bias corrections (corrected from IPP and rare event bias) both detailed in Annex A.3. Source: Quadros de Pessoal.

To both facilitate the decomposition and to be more computationally efficient (especially for the average partial effects), I create a D variable. This variable is the unique combination of the variables $teen_{it}$, mw_{it} and t. This modification does not change the results in any way, since the procedures are equivalent (Olden and Møen 2022). The α from model (26) is computed as follows: first, I take the age differences (teen minus young workers), then the differences of wage bins (MW minus above), and finally the difference between the years (year t minus 1986).

On the auxiliary regressions, the general auxiliary equations (6) becomes (28), (29), (30) and (31); the general CARs (9) become (32), (33), (34) and (35); the general RE (7) is (36) and the base and full are the aforementioned (27) and (26), respectively. To decompose the FE, I use Raposo, Portugal, and Carneiro (2019) method and create indicative matrices $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ in equations (30), (31).

Auxiliary Regressions:

$$\textbf{log(firm size)} = \boldsymbol{\Gamma}_0^{firm\ size} + \boldsymbol{D}\boldsymbol{\Gamma}^{firm\ size} + \boldsymbol{v}^{firm\ size} \tag{28}$$

$$\textbf{tenure} = \boldsymbol{\Gamma}_0^{tenure} + \boldsymbol{D}\boldsymbol{\Gamma}^{tenure} + \boldsymbol{v}^{tenure} \tag{29}$$

$$\boldsymbol{M}_1\hat{\boldsymbol{\theta}} = \boldsymbol{\Gamma}_0^{\theta} + \boldsymbol{D}\boldsymbol{\Gamma}^{\theta} + \boldsymbol{v}^{\theta} \tag{30}$$

$$\boldsymbol{M}_2\hat{\boldsymbol{\psi}} = \boldsymbol{\Gamma}_0^{\psi} + \boldsymbol{D}\boldsymbol{\Gamma}^{\psi} + \boldsymbol{v}^{\psi} \tag{31}$$

CARs:

$$\boldsymbol{v}^{firm\ size} = \boldsymbol{\delta}_0^{firm\ size} + \boldsymbol{D}\boldsymbol{\delta}_1^{firm\ size} + \boldsymbol{Y}\boldsymbol{\delta}_2^{firm\ size} + \boldsymbol{e}^{firm\ size} \tag{32}$$

$$\boldsymbol{v}^{tenure} = \boldsymbol{\delta}_0^{tenure} + \boldsymbol{D}\boldsymbol{\delta}_1^{tenure} + \boldsymbol{Y}\boldsymbol{\delta}_2^{tenure} + \boldsymbol{e}^{tenure} \tag{33}$$

$$\boldsymbol{v}^{\theta} = \boldsymbol{\delta}_0^{\theta} + \boldsymbol{D}\boldsymbol{\delta}_1^{\theta} + \boldsymbol{Y}\boldsymbol{\delta}_2^{\theta} + \boldsymbol{e}^{\theta} \tag{34}$$

$$\boldsymbol{v}^{\psi} = \boldsymbol{\delta}_0^{\psi} + \boldsymbol{D}\boldsymbol{\delta}_1^{\psi} + \boldsymbol{Y}\boldsymbol{\delta}_2^{\psi} + \boldsymbol{e}^{\psi} \tag{35}$$

RE:

$$ln\left(\frac{Pr(y_{it}=1|X)}{1-Pr(y_{it}=1|X)}\right) = \boldsymbol{D}\boldsymbol{\beta}_{RE} + \hat{\boldsymbol{v}}^{\theta}\boldsymbol{\Lambda}_1 + \hat{\boldsymbol{v}}^{\psi}\boldsymbol{\Lambda}_2 + \hat{\boldsymbol{v}}^{firm\ size}\boldsymbol{\Lambda}_3 + \hat{\boldsymbol{v}}^{tenmure}\boldsymbol{\Lambda}_4 + \boldsymbol{u}_{full} \tag{36}$$

Table 5: Coefficients for the Logit Decomposition

	Variables			**Coefficients Logit**		
	Year	Wage	Age	Base model	RE model	Full model
Bins	1986	MW	18-19	-0.01	-0.07	-0.88
	1986	MW	25-29	-0.13	-0.22	-0.85
	1986	Above	18-19	-0.14	-0.25	-1.00
	1986	Above	25-29	-0.33	-0.45	-0.82
	1988	MW	18-19	0.38	0.40	-0.19
	1988	MW	25-29	0.53	0.58	0.19
	1988	Above	18-19	0.48	0.51	-0.02
	1988	Above	25-29	0.19	0.18	0.03
Non-structural variables	Worker FE			No	No	Yes
	Firm FE			No	No	Yes
	log(firm size)			-	-	-0.022937
	tenure			-	-	-0.002258
	res. of firm of size			-	-0.022937	-
	res. of tenure			-	-0.002258	-
	res. of worker			-	0.999898	-
	res. of firm			-	0.999897	-

Notes: In this table, the D coefficients for the DDD and control variables are shown, from the updated sample, without any bias correction. The full 26, the base 27, and the RE model 36 are estimated with D and no interaction terms. All models are estimated with singletons and perfectly classified observations removed. The sample excludes workers from Madeira or Azores, in the primary sector, with a part-time job and independently employed. Source: Quadros de Pessoal.

Table 5 shows the results from the updated sample of the base, RE, and full models for the main D bins. All without bias corrections, because in this section I focus on building the basis for the decomposition. From base to RE we have rescaling effects and from RE to full there are the confounding effects, by each variable. As expected, $\widehat{\Lambda}_1 = \widehat{\Lambda}_2 = 1$ because dependent variables of the FE auxiliary equations (30) and (31) already have the estimated values, $\widehat{\Lambda}_3 = \widehat{\xi}_1$ and $\widehat{\Lambda}_4 = \widehat{\xi}_2$. I took the following steps to perform the logit decomposition shown in Table 6:

1. Estimate the TWFE logit Full model (26) with the D variable instead of the interaction terms, as explained in Section A.3.4;
2. Estimate the Base model (27) with the MLE, the D variable and a sample without singletons nor perfectly classified workers and firms;
3. Estimate the linear auxiliary regressions (28) (29) (30) (31) and keep both $\widehat{\Gamma}$'s and $\widehat{\upsilon}$'s;
4. Estimate the RE (36). Keep the $\widehat{\Lambda}$s;
5. Calculate $\widehat{correlated\ bias} = \widehat{\beta}_{RE} - \widehat{\beta}_{full} = \widehat{\Gamma}^{\theta} M_1 \widehat{\theta} + \widehat{\Gamma}^{\psi} M_1 \widehat{\psi} + \widehat{\Gamma}^{firm\ size} \widehat{\xi}_1 + \widehat{\Gamma}^{tenure} \widehat{\xi}_2$;
6. Estimate the linear auxiliary regressions (32) (33) (34) (35) and keep both $\widehat{\delta}_1$'s and $\widehat{\delta}_2$'s;
7. Estimate the precision matrix of the assumed multivariate conditionally normal distribution of $\widehat{\upsilon}$'s using the method of Yuan and Lin (2007) (details in Annex A.3.9). Keep the non-diagonal terms;
8. Calculate the diagonal terms of the precision matrix with the constraint on equation (23);
9. Calculate the bias conditional on the state of y:

$$\widehat{rescaling\ bias} = \widehat{\delta}_2^{\theta} \widehat{\delta}_1^{\theta} \widehat{\xi}_{\theta,\theta} + \widehat{\delta}_2^{\theta} (\widehat{\Lambda}^{\theta} - \widehat{\delta}_2^{\theta} \widehat{\xi}_{\theta,\theta}) + \tag{37}$$

$$\widehat{\delta}_2^{\psi} \widehat{\delta}_1^{\psi} \widehat{\xi}_{\psi,\psi} + \widehat{\delta}_2^{\psi} (\widehat{\Lambda}^{\psi} - \widehat{\delta}_2^{\psi} \widehat{\xi}_{\psi,\psi}) + \tag{38}$$

$$\widehat{\delta}_2^{ten} \widehat{\delta}_1^{ten} \widehat{\xi}_{ten,ten} + \widehat{\delta}_2^{ten} (\widehat{\Lambda}^{ten} - \widehat{\delta}_2^{ten} \widehat{\xi}_{ten,ten}) + \tag{39}$$

$$\widehat{\delta}_2^{size} \widehat{\delta}_1^{size} \widehat{\xi}_{size,size} + \widehat{\delta}_2^{size} (\widehat{\Lambda}^{size} - \widehat{\delta}_2^{size} \widehat{\xi}_{size,size}) \tag{40}$$

4.6 Decomposition Results

Table 6 shows the decomposed D coefficients of interest for the triple difference methodology, i.e., the relevant bins to identify the effect of the MW, shown in the row "minimum wage" (equivalent to the parameter α_{88} in the base and full models).

The decomposition is divided by confounding bias under the Gelbach part and rescaling bias under the CAR part. Notice that the results of the CAR decomposition add up closely to the actual bias conditional on the state of y, i.e., column (15) -0.09 is close to column (5) -0.10, implying a somewhat normal distribution of the residuals of the CAR equations (32), (33), (34) and (35). However, none of the residuals is normal, they all fail the Shapiro-Wilk test (although firm and worker

FE get quite close, with p-values near 0.06). But the residuals still resemble normality enough to enable an approximation of the decomposition. See Annex A.3.10 for some examples of qqplots and density distributions of the residuals of the auxiliary regressions, which show that $\hat{\boldsymbol{v}}^{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{v}}^{\boldsymbol{\psi}}$ are *almost* normal distributions and tenure $\hat{\boldsymbol{v}}^{\boldsymbol{tenure}}$ and log(firm size) $\hat{\boldsymbol{v}}^{\boldsymbol{firm\ size}}$ residuals are closer to log-normal distributions.

Interpretation of the Decomposition

The interpretation of Table 6 is not direct. The rows are bins/group of workers and the row "minimum wage" is the triple difference methodology that identifies the MW effect from those bins. Consider the following insights.

Firstly, the base coefficients in column (1) are the logit *log-odds* equivalent to that group's raw separation rate. Secondly, a positive bias from a control variable implies that its inclusion diminishes that bin's attributable part of the separation propensity; the control variable is responsible for that share of separation propensity.

Thus, when I use Gelbach (2009) decomposition to decompose the MW displacement effect, I get the differences in the incidence of the MW (that would be biasing the analysis). For example, omitting worker constant characteristics leads to a smaller MW effect of -0.10, because the MW affects less productive workers.

Several conclusions can be drawn from the Gelbach part. First, in columns (6) and (7) we see that both workers and firms had a big role in accounting for the separation rate of a bin. However, the firm had consistently the highest bias on all bins, making it the most important determinant of separations, as seen in column (7) (which makes sense, since separations are mostly lay-offs and hardly any quits (Davis, Faberman, and Haltiwanger 2012)). Secondly, in 1988, it's possible to see that teens have similar shares of separations explained by tenure, worker and firm FEs, suggesting that teens above and at the MW are employed in similar firms, have similar tenure and similar individual propensity, which validates my assumption of their proximity to control for teen trends. This assumption has been made before in the MW analysis of Abowd et al. (1997). The positive contribution of tenure in the confounding bias is also coherent with the empirical fact that workers who are hired last, are the first to leave the firm (Buhai et al. 2014).

Table 6: Logit Decomposition of the DDD specification.

Bins			Model's Coeffs			True Bias		Decomposition										
								Confounding Effects: Gelbach					Rescaling Effects: CAR					Total
Year	Wage	Age	Base (1)	RE (2)	Full (3)	Total (4)	Uncorr. (5)	Worker (6)	Firm (7)	Size (8)	Tenure (9)	**Sum** (10)	Worker (11)	Firm (12)	Size (13)	Tenure (14)	**Sum** (15)	(16)
1986	MW	18-19	-0.01	-0.07	-0.88	0.86	0.05	0.00	0.63	0.02	0.16	0.81	0.01	0.02	0.00	0.01	0.04	0.85
1986	MW	25-29	-0.13	-0.22	-0.85	0.71	0.08	-0.01	0.60	-0.00	0.04	0.63	0.02	0.04	0.00	0.01	0.07	0.70
1986	Above	18-19	-0.14	-0.25	-1.00	0.86	0.11	0.08	0.53	-0.00	0.14	0.75	0.02	0.04	0.01	0.02	0.09	0.84
1986	Above	25-29	-0.33	-0.45	-0.82	0.49	0.12	-0.03	0.37	-0.03	0.06	0.37	0.02	0.06	0.01	0.02	0.11	0.48
1988	MW	18-19	0.38	0.40	-0.19	0.57	-0.01	-0.10	0.52	0.00	0.16	0.58	-0.01	-0.01	-0.00	-0.00	-0.03	0.56
1988	MW	25-29	0.53	0.58	0.19	0.33	-0.06	-0.27	0.54	0.01	0.10	0.39	-0.01	-0.05	-0.00	-0.01	-0.07	0.32
1988	Above	18-19	0.48	0.51	-0.02	0.50	-0.04	-0.08	0.47	-0.00	0.15	0.54	-0.01	-0.03	-0.00	-0.00	-0.04	0.50
1988	Above	25-29	0.19	0.18	0.03	0.16	0.01	-0.25	0.34	-0.02	0.08	0.15	0.01	-0.01	-0.00	-0.00	-0.00	0.15
minimum wage			0.36	0.46	0.48	−0.12	−0.10	−0.10	0.02	0.01	0.05	−0.02	−0.03	−0.06	0.00	−0.01	−0.09	−0.11
Perc. of total bias						(100%)	(83%)	(81%)	(−17%)	(−7%)	(−41%)	(17%)	(22%)	(51%)	(0%)	(6%)	(79%)	(95%)
Perc. of est. CAR bias													(28%)	(65%)	(0%)	(7%)	(100%)	

Notes: Columns (1), (2), and (3) are the model's coefficients without any bias correction, from models 27, 36, and 26, respectively. The total bias is column $(4) = (1) - (3)$. The uncorrelated component of the bias is $(5) = (1) - (2)$. From (6) to (9) are the variables' components of the Gelbach decomposition applied to the logit model; (10) is its total. From (11) to (14) are the variables' components of the CAR applied to the logit model; (15) is its total. $(16) = (10) + (15)$. The DDD line is the DDD estimate decomposed, calculated as: $[(1988\ MW\ teen - 1988\ MW\ young) - (1988\ Above\ teen - 1988\ Above\ young)] - [(1986\ MW\ teen - 1986\ MW\ young) - (1986\ Above\ teen - 1986\ Above\ young)]$. The line Perc. of total bias are percentage values computed in relation to column (4). The last line, Perc. of est. CAR bias, are percentages of the estimated total CAR bias from column (15). Source: Quadros de Pessoal.

By using the CAR decomposition, we can see the main determinants of the MW impact on the probability of separation. The CAR divides the heterogeneity in the MW displacement effects by measuring how essential each control variable was for the materialization of the logit *log-odds* in actual separations in the data.

The heterogeneity is explained as follows: 65% by the firm, 28% by the worker, 7% by tenure, and almost 0% by firm size. This leads me to conclude that the most important factor to determine if the effect of the MW in separations is higher or lower is the firm where that worker is employed at.[23]

5 Issues and Caveats

In this section, I acknowledge the main downsides of this thesis.

Firstly, although I create a counterfactual that accounts for shocks to teens and to MW workers, if in 1986 there is any other shock that is specific to the separations of MW teen workers, in my setting it's impossible to disentangle it from the MW effects.

Secondly, to include the worker and firm FEs (which warrant validity to *Assumption 1* and allow the decomposition of the MW effect by worker and firm roles for Section 4.6), there is an increase, albeit not statistically significant, in the coefficient of interest α. In Table 3 we can see that the number of singletons that separate diminishes considerably and is inconsistent throughout the bins, because some bins have more observations from previous years. This asymmetry of deletes is a caveat of this specification, that increases the estimates, biasing them slightly.

Thirdly, there is selective group bias. Teens are regarded as less resilient workers, making them potentially more vulnerable to MW increases and not representative of the entire population (Card 1992; Neumark and Wascher 1995b; Allegretto, Dube, and Reich 2011).

Fourthly, a careful estimation of an AKM (Abowd, Kramarz, and Margolis 1999) model should be made on (at least!) the largest connected set (Abowd, Creecy, and Kramarz 2002). More recent literature even recommends econometricians go further, and estimate a leave-one-out set (Kline, Saggio, and Sølvsten 2020). Unfortunately, I cannot do it because computing the largest connected set would eliminate observations that are critical to the identification of the MW effect. Thus, I stick with the simplest method and just delete singletons and perfectly classified.

Finally, there are issues in the decomposition. I lack both control variables that follow an exact normal distribution and formal proof of the interpretation of the rescaling effects.

[23] Furthermore, I do not confirm the results of some previous literature, like Rama (2001), which has found that the higher the firm size, the smaller the MW effects, because I find that firm size is not a determinant of the MW effect.

6 Conclusion

In this thesis, I develop a novel conditional decomposition for the logit model. The decomposition divides the impact of covariates on structural parameters. Because of a rescaling feature of the logit model, the decomposition is capable of dividing the importance of each variable at explaining the heterogeneity of a treatment effect. I apply the decomposition to a minimum wage natural experiment in 1987 Portugal. The impact of the minimum wage was heterogeneous, and those differences between workers are explained in 65% by the firm where a worker is employed, 28% by the worker, and 7% by tenure.

The decomposition can be helpful in future applications. For policy evaluation, it allows the researcher to detect which factors exacerbate the nefarious/beneficial effects of a policy. With that information, policymakers could discourage/incentive those factors, helping programs to target better the treatment group. Any binary problem, like unemployment or being pregnant, can be studied with a logit and can use the theory here developed.

Notwithstanding, further research is needed. Firstly, the developed decomposition is only an approximation, which assumes conditional normality of the residuals from regressing the control variables on the structural variables. This may be improved by generalizing the procedure using a more malleable distribution, like the Weibull. Secondly, a variance-covariance matrix of the decomposition's components must be developed. Thirdly, I provided no formal mathematical proof for the interpretation of the rescaling effects. Lastly, a natural extension would be applying the theory developed in this thesis to the multinomial logit case.

7 References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey Wooldridge. 2017. "When Should You Adjust Standard Errors for Clustering?" Working Paper 24003. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w24003.

Abdul-Hamid, H, and J P Nolan. 1998. "Multivariate Stable Densities as Functions of Their One Dimensional Projections." *Multivariate Analysis* 67: 80–89.

Abowd, John, Robert Creecy, and Francis Kramarz. 2002. "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data." Longitudinal Employer-Household Dynamics Technical Papers 2002-06. Center for Economic Studies, U.S. Census Bureau.

Abowd, John, Francis Kramarz, Thomas Lemieux, and David N. Margolis. 1997. "Minimum Wages and Youth Employment in France and the United States." *NBER Working Paper*, no. w6111.

Abowd, John, Francis Kramarz, and David N Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica* 67 (2): 251–333.

Addison, J. T., P. Portugal, and H. Vilares. 2017. "Unions and Collective Bargaining in the Wake of the Great Recession : Evidence from Portugal." *British Journal of Industrial Relations.* 55 (3): 551–76.

Allegretto, Sylvia, Arindrajit Dube, and Michael Reich. 2011. "Do Minimum Wages Really Reduce Teen Employment? Accounting for Heterogeneity and Selectivity in State Panel Data." *Industrial Relations: A Journal of Economy and Society* 50 (2): 205–40.

Allegretto, Sylvia, Arindrajit Dube, Michael Reich, and Ben Zipperer. 2017. "Credible Research Designs for Minimum Wage Studies: A Response to Neumark, Salas, and Wascher." *ILR Review* 70 (3): 559–92. https://doi.org/10.1177/0019793917692788.

Allison, Paul D. 1999. "Comparing Logit and Probit Coefficients Across Groups." *Sociological Methods & Research* 28 (2): 186–208. https://doi.org/10.1177/0049124199028002003.

Amemiya, T. 1985. *Advanced Econometrics*. Harvard Univ. Press.

Banerjee, Onureena, Laurent El Ghaoui, Alexandre d'Aspremont, and Georges Natsoulis. 2006. *Convex Optimization Techniques for Fitting Sparse Gaussian Graphical Models*. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/1143844.1143856.

Boockmann, Bernhard. 2010. "The Combined Employment Effects of Minimum Wages and Labor Market Regulation – a Meta-Analysis." IAW Discussion Papers, no. 65 (May).

Bossler, Mario, and Hans-Dieter Gerner. 2020. "Employment Effects of the New German Minimum Wage: Evidence from Establishment-Level Microdata." *ILR Review* 73 (5): 1070–94. https://doi.org/10.1177/0019793919889635.

Botosaru, Irene, and Chris Muris. 2017. "Binarization for Panel Models with Fixed Effects."

Bound, John, and Richard B. Freeman. 1992. "What Went Wrong? The Erosion of Relative Earnings and Employment Among Young Black Men in the 1980s*." *The Quarterly Journal of Economics* 107 (1): 201–32. https://doi.org/10.2307/2118327.

Brown, Charles, Curtis Gilroy, and Andrew Kohen. 1982. "The Effect of the Minimum Wage on Employment and Unemployment: A Survey," Working paper series, no. 846. https://doi.org/10.3386/w0846.

Buckles, Kasey S., and Daniel M. Hungerman. 2013. "Season of Birth and Later Outcomes: Old Questions, New Answers." *The Review of Economics and Statistics* 95 (3): 711–24. https://doi.org/10.1162/REST_a_00314.

Buhai, I. Sebastian, Miguel A. Portela, Coen N. Teulings, and Aico van Vuuren. 2014. "Returns to Tenure or Seniority?" *Econometrica* 82 (2): 705–30. https://doi.org/https://doi.org/10.3982/ECTA8688.

Burkhauser, Richard V., Kenneth A. Couch, and David C. Wittenburg. 1996. "”Who Gets What” from Minimum Wage Hikes: A Re-Estimation of Card and Krueger's Distributional Analysis in Myth and Measurement: The New Economics of the Minimum Wage." *Industrial and Labor Relations Review* 49 (3): 547–52.

Card, David. 1992. "Do Minimum Wages Reduce Employment? A Case Study of California, 1987–89." *ILR Review* 46 (1): 38–54. https://doi.org/10.1177/001979399204600104.

Card, David, Ana Rute Cardoso, and Patrick Kline. 2015.“ Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women *." *The Quarterly Journal of Economics* 131 (2): 633–86. https://doi.org/10.1093/qje/qjv038.

Card, David, and Alan B Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *American Economic Review* 84 (4): 772–93.

Card, David, and Alan B. Krueger. 1995. *Myth and Measurement: The New Economics of the Minimum Wage*. Economics Books 10738. Princeton University Press.

Carruthers, Celeste K., and Marianne H. Wanamaker. 2017. "Separate and Unequal in the Labor Market: Human Capital and the Jim Crow Wage Gap." *Journal of Labor Economics* 35 (3): 655–96. https://doi.org/10.1086/690944.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer. 2019. "The Effect of Minimum Wages on Low-Wage Jobs." *The Quarterly Journal of Economics* 134 (3): 1405–54.

Cerejeira, João. 2008. "Young Employment, Job-Skill Composition and Minimum Wages: Evidence from a 'Natural Experiment'." NIPE Working Papers, no. 04/2008. https://ideas.repec.org/p/nip/nipewp/04-2008.html.

Charles, Kerwin Kofi, and Jonathan Guryan. 2008. "Prejudice and Wages: An Empirical Assessment of Becker's the Economics of Discrimination." *Journal of Political Economy* 116 (5):

773–809. https://doi.org/10.1086/593073.

Cramer, J. S. 2007. "Robustness of Logit Analysis: Unobserved Heterogeneity and Mis-Specified Disturbances*." *Oxford Bulletin of Economics and Statistics* 69 (4): 545–55.

Cruz-Gonzalez, Mario, Iván Fernández-Val, and Martin Weidner. 2017. "Bias Corrections for Probit and Logit Models with Two-Way Fixed Effects." *The Stata Journal* 17 (3): 517–45. https://doi.org/10.1177/1536867X1701700301.

Czarnowske, Daniel, and Amrei Stammann. 2019. "Fixed Effects Binary Choice Models: Estimation and Inference with Long Panels."

Davis, Steven J., R. Jason Faberman, and John Haltiwanger. 2012. "Labor Market Flows in the Cross Section and over Time." *Journal of Monetary Economics* 59 (1): 1–18. https://doi.org/https://doi.org/10.1016/j.jmoneco.2011.10.001.

Davison, A. C., and D. V. Hinkley. 1997. *Bootstrap Methods and Their Application*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press. https://doi.org/10.1017/CBO9780511802843.

Doeringer, P. B., and M. J. Piore. 1971. "Internal Labor Markets and Manpower Analysis: With a New Introduction." In, 1st ed. Routledge.

Dolado, Juan J., Francis Kramarz, Stephen J. Machin, Alan Manning, David N. Margolis, and Coen N. Teulings. 1996. "The Economic Impact of Minimum Wages in Europe." *Economic Policy* 11: 317–72.

Dube, Arindrajit. 2018. "Impacts of minimum wages: review of the international evidence." 978-1-912809-89-9. Her Majesty's Treasury: Department for Business, Energy & Industrial Strategy.

Dube, Arindrajit, T. William Lester, and Michael Reich. 2010. "Minimum Wage Effects Across State Borders: Estimates Using Contiguous Counties." *The Review of Economics and Statistics* 92 (4): 945–64.

Dubin, Jeffrey, and Douglas Rivers. 1989. "Selection Bias in Linear Regression, Logit and Probit Models." *Sociological Methods & Research* 18 (2-3): 360–90. https://doi.org/10.1177/0049124189018002006.

Fan, Jianqing, Yuan Liao, and Han Liu. 2016. "An overview of the estimation of large covariance and precision matrices." *The Econometrics Journal* 19 (1): C1–32. https://doi.org/10.1111/ectj.12061.

Fernández-Val, Ivan, and Martin Weidner. 2016. "Individual and Time Effects in Nonlinear Panel Models with Large n, t." *Journal of Econometrics* 192 (1): 291–312.

Fernández-Val, Iván, and Martin Weidner. 2018. "Fixed Effect Estimation of Large t Panel Data Models."

Fisher, R. A. 1935. *The Design of Experiments*. Oliver & Boyd.

Fortin, Nicole, Thomas Lemieux, and Sergio Firpo. 2011. "Chapter 1 - Decomposition Methods in Economics." In, edited by Orley Ashenfelter and David Card, 4:1–102. Handbook of Labor Economics. Elsevier. https://doi.org/https://doi.org/10.1016/S0169-7218(11)00407-2.

Gaure, S. 2013. "OLS with Multiple High Dimensional Category Variables." *Computational Statistics and Data Analysis* 66: 8–18.

Gelbach, Jonah B. 2009. "When Do Covariates Matter? And Which Ones, and How Much." *Journal of Labour Economics*.

Glymour, Madelyn, Judea Pearl, and Nicholas P Jewell. 2016. *Causal Inference in Statistics: A Primer*. John Wiley & Sons.

Golberger, Arthur. 1991. *A Course in Econometrics*. Harvard Univ. Press.

Gregory, Terry, and Ulrich Zierahn. 2020. "When the Minimum Wage Really Bites Hard: Impact on Top Earners and Skill Supply." CESifo Working Paper Series, no. 8540.

Hahn, Jinyong. 1995. "Bootstrapping Quantile Regression Estimators." *Econometric Theory* 11 (1): 105–21. https://doi.org/10.1017/S0266466600009051.

Halaby, Charles N. 2004. "Panel Models in Sociological Research: Theory into Practice." *Annual Review of Sociology* 30 (1): 507–44. https://doi.org/10.1146/annurev.soc.30.012703.110629.

Harasztosi, Peter, and Attila Lindner. 2019. "Who Pays for the Minimum Wage?" *American Economic Review* 109 (8): 2693–2727.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47 (1): 153–61.

Hellerstein, Judith K., and David Neumark. 2008. "Workplace Segregation in the United States: Race, Ethnicity, and Skill." *The Review of Economics and Statistics* 90 (3): 459–77. https://doi.org/10.1162/rest.90.3.459.

Jardim, Ekaterina, Mark C. Long, Robert Plotnick, Emma van Inwegen, Jacob Vigdor, and Hilary Wething. 2022. "Minimum-Wage Increases and Low-Wage Employment: Evidence from Seattle." *American Economic Journal: Economic Policy* 14 (2): 263–314. https://doi.org/10.1257/pol.20180578.

Kaitz, Hyman. 1970. "Experience of the Past: The National Minimum." *Youth Unemployment and Minimum Wages*, 30–54.

Karlson, Kristian Bernt, Anders Holm, and Richard Breen. 2012. "Comparing Regression Coefficients Between Same-Sample Nested Models Using Logit and Probit: A New Method." *Sociological Methodology* 42 (1): 286–313. https://doi.org/10.1177/0081175012444861.

King, Gary, and Langche Zeng. 2001. "Logistic Regression in Rare Events Data." *Political Analysis* 9.

Kline, Patrick, Raffaele Saggio, and Mikkel Sølvsten. 2020. "Leave-Out Estimation of Variance Components." *Econometrica* 88 (5): 1859–98.

Lancaster, Tony. 2000. "The Incidental Parameter Problem Since 1948." *Journal of Econometrics*

95 (2): 391–413. https://doi.org/https://doi.org/10.1016/S0304-4076(99)00044-5.

Lauritzen, Steffen L. 1996. *Graphical Models*. *Oxford University Press*.

Lee, Kyoungjae, and Jaeyong Lee. 2017. "Estimating Large Precision Matrices via Modified Cholesky Decomposition." *arXiv: Statistics Theory*.

Lee, Lung. 1980. "Specification Error in Multinomial Logit Models: Analysis of the Omitted Variable Bias," Discussion paper, center for economic research, department of economics, no. 131.

———. 1982. "Specification Error in Multinomial Logit Models: Analysis of the Omitted Variable Bias." *Journal of Econometrics* 20 (2): 197–209.

MaCurdy, Thomas. 2015. "How Effective Is the Minimum Wage at Supporting the Poor?" *Journal of Political Economy* 123 (2): 497–545.

Martins, Pedro. 2014. "30,000 Minimum Wages: The Economic Effects of Collective Agreement Extensions." Working Papers 51. Queen Mary, University of London, School of Business; Management, Centre for Globalisation Research.

Mood, Carina. 2009. "Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It." *European Sociological Review* 26 (1): 67–82.

———. 2017. "Logistic Regression: Uncovering Unobserved Heterogeneity." *Unpublished Manuscript. Https://Www. Su. Se/Polopoly_fs/1.341160* 1501927239.

Neumark, David, J. M. Ian Salas, and William Wascher. 2014. "Revisiting the Minimum Wage—Employment Debate: Throwing Out the Baby with the Bathwater?" *ILR Review* 67 (3_suppl): 608–48. https://doi.org/10.1177/00197939140670S307.

Neumark, David, Mark Schweitzer, and William Wascher. 2000. "The Effects of Minimum Wages Throughout the Wage Distribution." NBER Working Papers, no. 7519 (February).

Neumark, David, and Peter Shirley. 2021. "Myth or Measurement: What Does the New Minimum Wage Research Say about Minimum Wages and Job Loss in the United States?" Working Paper, Working paper series, no. 28388.

Neumark, David, and William Wascher. 1995a. "Minimum Wage Effects on Employment and School Enrollment." *Journal of Business & Economic Statistics* 13 (2): 199–206.

———. 1995b. "Minimum-Wage Effects on School and Work Transitions of Teenagers." *American Economic Review* 85 (2): 244–49.

———. 2000. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Comment." *The American Economic Review* 90 (5): 1362–96.

———. 2004. "Minimum Wages, Labor Market Institutions, and Youth Employment: A Cross-National Analysis." *ILR Review* 57 (2): 223–48. https://doi.org/10.1177/001979390405700204.

———. 2006. "Minimum Wages and Employment: A Review of Evidence from the New Minimum Wage Research." Working Paper, Working paper series, no. 12663. https://doi.org/10.3386/w12663.

Neyman, J., and Elizabeth L. Scott. 1948. "Consistent Estimates Based on Partially Consistent Observations." *Econometrica* 16 (1): 1–32.

Olden, Andreas, and Jarle Møen. 2022. "The triple difference estimator." *The Econometrics Journal*, March. https://doi.org/10.1093/ectj/utac010.

Pereira, Sonia. 2003. "The Impact of Minimum Wages on Youth Employment in Portugal." *European Economic Review* 47 (April): 229–44. https://doi.org/10.1016/S0014-2921(02)00209-X.

Phillips, Peter C. B., and Hyungsik R. Moon. 1999. "Linear Regression Limit Theory for Nonstationary Panel Data." *Econometrica* 67 (5): 1057–1111. https://doi.org/https://doi.org/10.1111/1468-0262.00070.

Portugal, Pedro, and Ana Rute Cardoso. 2006. "Disentangling the Minimum Wage Puzzle: An Analysis of Worker Accessions and Separations." *Journal of the European Economic Association* 4 (5): 988–1013. https://doi.org/10.1162/JEEA.2006.4.5.988.

Rama, Martin. 2001. "The Consequences of Doubling the Minimum Wage: The Case of Indonesia." *Industrial and Labor Relations Review* 54 (4): 864–81.

Raposo, Pedro, Pedro Portugal, and Anabela Carneiro. 2019. "The Sources of the Wage Losses of Displaced Workers: The Role of the Reallocation of Workers into Firms, Matches, and Job Titles." *Journal of Human Resources* 56 (September): 0317–8667R3. https://doi.org/10.3368/jhr.56.3.0317-8667R3.

Roth, Jonathan, Pedro H. C. Sant'Anna, Alyssa Bilinski, and John Poe. 2022. "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature." https://doi.org/10.48550/ARXIV.2201.01194.

Stammann, Amrei. 2018. "Fast and Feasible Estimation of Generalized Linear Models with High-Dimensional k-Way Fixed Effects." https://arxiv.org/abs/1707.01815.

Stewart, Mark B. 2002. "Estimating the Impact of the Minimum Wage Using Geographical Wage Variation*." *Oxford Bulletin of Economics and Statistics* 64 (supplement): 583–605.

Stigler, George J. 1946. "The Economics of Minimum Wage Legislation." *The American Economic Review* 36 (3): 358–65.

Train, Kenneth E. 2009. "Logit." In *Discrete Choice Methods with Simulation*, 2nd ed., 34–75. Cambridge University Press. https://doi.org/10.1017/CBO9780511805271.003.

Williams, Richard. 2009. "Using Heterogeneous Choice Models to Compare Logit and Probit Coefficients Across Groups." *Sociological Methods & Research* 37 (4): 531–59. https://doi.org/10.1177/0049124109335735.

Winship, Christopher, and Robert D. Mare. 1984. "Regression Models with Ordinal Varibles." *American Sociological Review* 49: 512.

Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. MIT Press.

Yuan, Ming, and Yi Lin. 2007. "Model selection and estimation in the Gaussian graphical model." *Biometrika* 94 (1): 19–35. https://doi.org/10.1093/biomet/asm018.

A Annexes

A.1 An R Package for the Logit Decomposition

I built an R package to implement the novel logit decomposition. It's called "Decomp", and it's ready to be deployed through *github*. The package has only one function, the `logit_decomposition()`, which is pipe friendly. To download it, type in R:

```
devtools::install_github("Salema-DG/Decomp")
```

Basic Documentation

Usage:

```
logit_decomposition(data,
                    dependent_var,
                    main_indep_var,
                    controls,
                    independent = FALSE)
```

Arguments:

- `data`: a data frame, data frame extension (e.g. a tibble), or a lazy data frame (e.g. from dbplyr or dtplyr);
- `dependent_var`: the binary dependent variable as an unquoted expression;
- `main_indep_var`: the main independent variables as an unquoted expression or a character vector in case of multiple variables;
- `controls`: the control variables as one or more unquoted expressions separated by commas;
- `independent`: logical argument. `TRUE` if the user wished to assume independence between the control variables, usign equation (12) for the decomposition.

An Example

Take `df1` as a tibble with 6 variables: `y` is a binary variable only with 0's and 1's and `x, z1, z2, z3, z4` are normally distributed variables. To perform the decomposition of the bias between a base model with only `x` as a dependent variable and a full model with `x, z1, z2, z3, z4`, we can type:

```
df1 %>% logit_decomposition(y, x, z1, z2, z3, z4, independent = TRUE)
```

The output on the R console is in Figure 7. The first column indicates the variable, the second column is the confounding effects (the part decomposable using Gelbach (2009)), the third column is the rescaling effects, and the last column is the sum of columns two and three. Row 6 is the estimated bias, which is the coefficient of x from the base model minus the coefficient of x from the full model. Rows 1 to 4 show the decomposition of the bias by each z variable. Row 5 sums all variables (rows 1 to 4), and this is the bias explained by the logit decomposition. Notice the first part is completely explained while the second is an approximation.

Figure 7: Output of the Decomp package

Decomposition of logit model

	variable	corr_part	uncorr_part	sum
1	z1	0.099	-0.0044	0.094
2	z2	0.078	-0.0077	0.070
3	z3	0.052	-0.0092	0.043
4	z4	0.017	-0.0060	0.011
5	sum	0.246	-0.0274	0.219
6	Biases	0.246	-0.0277	0.218

A.2 Data Annex

Quadros de Pessoal (QP) is a linked-employer-employee-dataset annually collected by the Ministry of Employment, Solidarity and Social Security (MTSSS). QP is collected with a mandatory survey to all private firms with at least one employee in Portugal. QP assembles information at the establishment, firm and worker level, with a fictitious ID for each of them. Worker's information is reported for personnel working at the reference week of QP, which was in March until 1993 and in October in the following years. Workers on short leave are included but workers on long leave are not; meaning, essentially, that workers on holidays, maternity/paternity leave, strikes, or sickness are included, but workers in the mandatory military service, which was in place until 2004, weren't reported.

Its legally mandatory nature ensures high response rates. Additionally, the Ministry's inspectors and the obligation to post the map of wages of the establishment in a public space of that establishment ensures adherence to the MW and collective agreements and the reliability of the information. Portuguese labour laws mandated the public showing of Quadros de Pessoal. But after the revision of the "Código de Trabalho" in 2009, it was revoked. Source: Diário da República, subparagraph r) of n.º 6 of article 12.º of the Law n.º 7/2009 and article 32.º of the Lei n.º 105/2009. Nowadays, QP collects information about more than 300 thousand firms and almost 3 million workers. Reported data includes: the worker's gender, earnings, hours worked, age, occupation, schooling, admission data, collective agreement coverage; firm's industry, size, sales, ownership, legal setting; location of the establishment, and location of the firm's headquarters.

Restrictions to the sample for this thesis

This thesis's sample was built with four main restrictions on QP:

- Eliminate all employees working in Madeira or Azores because the MW of these autonomous areas is set by their local parliaments. Both Madeira and Azores have a higher MW than mainland Portugal.
- Exclude agriculture, fishing, and pisciculture (from the primary sector only mining is kept). These sectors are excluded because of the low reliability of their data, due to the high informality and seasonality of labour.
- Keep only full-time employees, who are identified directly with a variable about the type of remuneration, and not by the hours worked in the reference month.
- Keep only dependent employed workers, who are identified directly with a variable about their professional situation.

For the estimation of the full equation (26), I further restrict the sample with:

- Eliminate singletons, i.e., workers or firms which only appeared once in QP;
- Eliminate perfectly classified, i.e., workers that always separate or never separate, and also firms that have workers that always separate or never separate.

Description of the variables

- Age (in years). I apply a max mode to the implied sequence of age, to diminish data entry errors. For example, if a worker is reported to be 40 years old in 1986, 41 in 1987, 40 in 1988, and 43 in 1989, I set the age in 1988 to 42 years old;
- Worker ID. It results from a transformation of the fiscal identifier of each Portuguese citizen. I apply the algorithm in Annex A.2.1 to select between duplicates;
- Firm ID. Each firm entering QP is assigned a unique identifier number. However, in 2010, a firm could be assigned a number of a "dead" firm, i.e., the firm ID of firms that closed before 2010 could be recycled to new firms. To partly solve this issue, I reassign the ID of firms created in 2010 that had the same number as firms that had closed already. Furthermore, if a firm closes in 2009 and its firm ID is recycled in 2010, there is no way to identify the closure of the firm with the previous method, so I use the founding year to tackle this issue.

- Tenure (in years). Just like to age, I apply a max mode to the implied sequence of tenure;
- Firm size. Results from a worker count within each firm after all the restrictions are applied;
- Year. The year of the reference month of QP.
- Wages. This variable is used to identify a minimum-wage and an above-minimum-wage worker. QP reports several monthly earnings. A researcher can have access to regularly paid base wage, payments for overtime, regularly paid subsidies (the biggest one is meal subsidy), extra payment for tenure, and irregular benefits. The MW is binding to the base wage. Using a full-time equivalence to adapt the base wage to the base hours worked (which are also reported in QP), I use a 5€ bandwidth to determine if a worker is or not earning the MW.
- Separations. A worker is considered to have separated from a firm if in the following year he is either outside of QP (either unemployed, self-employed or public servant) or if he is still in QP but working in a different firm.
- CPI. Because QP was collected in March until 1993 and in October henceforth, the inflation at the time of collection does not match the yearly reported inflation. Thus, I compute from monthly CPI retorts an annualized version taking into account the reference month.

A.2.1 The Algorithm for the Worker's ID

The QP dataset allows a worker to be reported under two or more different firms simultaneously if an individual has several jobs. It is standard practice to delete these duplicates (all the Portuguese papers mentioned in Section 4 *Literature* part do it). However, this approach blunderingly classifies separations, by assuming that workers that accept another job separate from their current one. Some Portuguese literature solves this with an algorithm that selects a primary firm to keep, using hours worked and/or wages.

However, my main goal is to avoid blunderingly classifying separations while also picking the *primary* firm. To achieve it, I prioritize continuity, always minimizing the number of separations, and using worker characteristics only to disentangle ambiguous situations. Figure 8 shows the saved observations using the algorithm, compared with complete duplicate removal; where it can be seen that the saved observations are quite common in the period of my analysis. Table 7 has all the possible situations of duplicates, and the following list refers to them.[24]

The algorithm:

1. If the duplicates are precisely equal, keep only one;
2. If the firm is the same, sequentially prefer:

[24] There are also cases, albeit less frequent, of a worker being reported twice or more by the same firm, which I treat as mistakes.

Table 7: Example of the Algorithm for Duplicates

(a) 1 before and after

Worker	Year	Firm
1	2010	100
1	2011	100
1	2011	101
1	2012	100
1	2013	100

(b) 1 before

Worker	Year	Firm
2	2010	100
2	2011	100
2	2011	101
2	2012	102
2	2013	102

(c) 1 after

Worker	Year	Firm
3	2010	100
3	2011	101
3	2011	102
3	2012	102
3	2013	102

(d) 2 before and 1 after

Worker	Year	Firm
4	2010	100
4	2010	101
4	2011	100
4	2011	101
4	2012	101
4	2013	101

(e) 1 before and 2 after

Worker	Year	Firm
5	2009	100
5	2010	100
5	2011	100
5	2011	101
5	2012	100
5	2012	101

(f) 2 before and after

Worker	Year	Firm
6	2010	100
6	2010	101
6	2011	100
6	2011	101
6	2012	100
6	2012	101

(g) 1 before and 1 after

Worker	Year	Firm
7	2009	100
7	2010	100
7	2010	101
7	2011	101

(h) None

Worker	Year	Firm
8	2009	102
8	2010	100
8	2010	100
8	2011	101

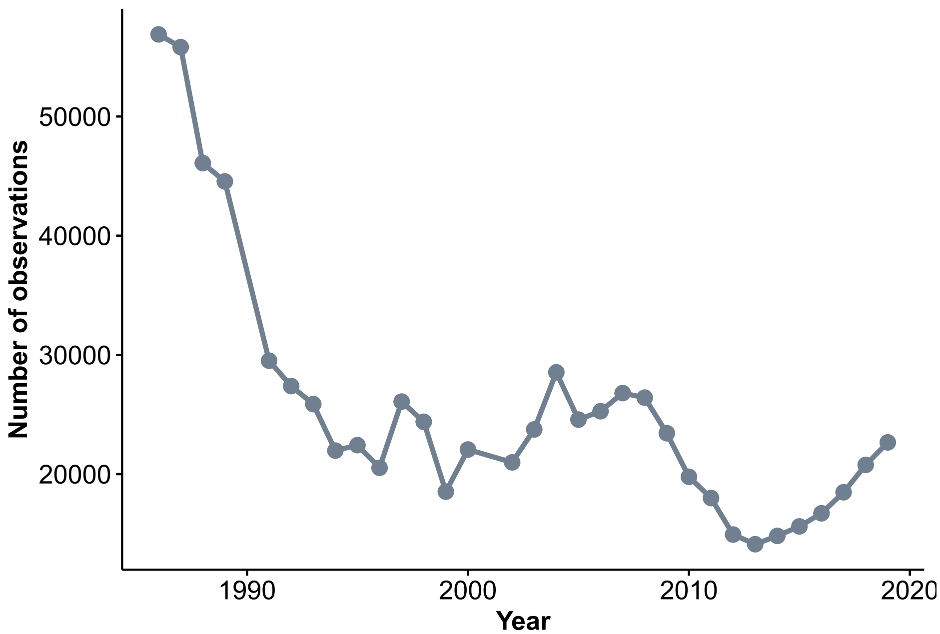
- The higher base wage;
- The higher normal hours worked;

3. If the firm is different, as listed in Table 1, produce the following feedback loop type of iteration (achieved with the `while` command in *R*) that minimizes separations:
   - Choose sole matches on (a), (b) and (c);
   - Choose the continuing match on (d) and (e);
   - If (f) or (g), repeat the previous steps. (f) may become a (c) or (d), and (g) may become a (b) or (c). Echo until the number of observations for worker i stabilizes. If (f) or (g) endures the feedback loop, move to step 4;
   - If (h) go to step 4;
4. Criteria to cast primary firm:
   - The firm with more appearances throughout the years;

- The one with the higher base wage;
- The higher normal hours worked;

5. To the detritus duplicates of the algorithm, pick randomly.

Figure 8: Algorithm's Saved Observations



Notes: This Figure shows the number of observations saved by the algorithm for the selection of duplicates in comparison with removing all worker duplicates. A worker duplicate is a worker that holds more than one job at the same time. Years from 1986 to 2019. Source: Quadros de Pessoal.

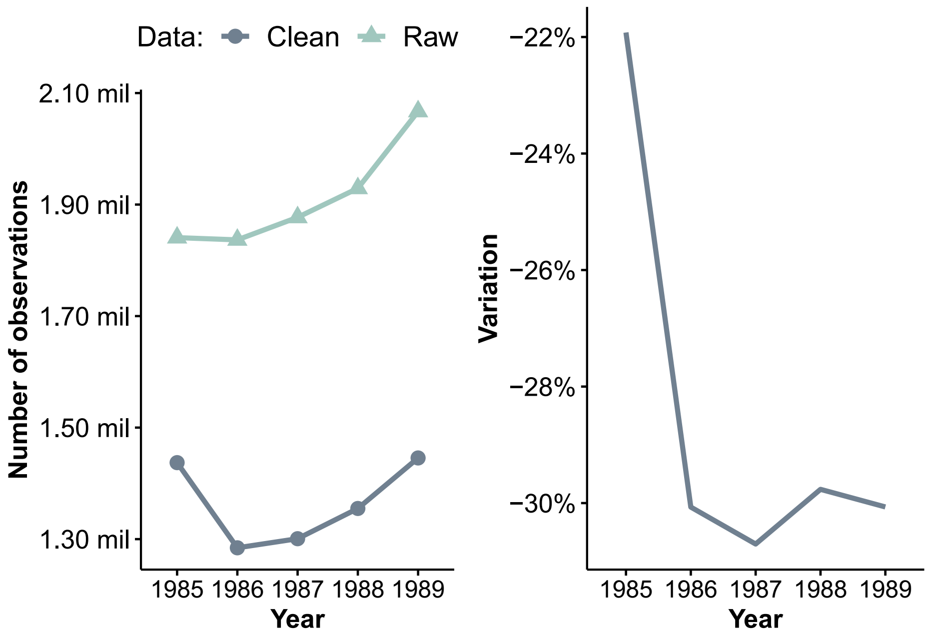
A.2.2 Problems with the year 1985

The aforementioned data algorithm in Section A.2.1 cannot be used for 1985 data because this year doesn't have an available worker ID. Subsequently, 1985 will have more observations than workers, an *overcounting*. However, in the following years, workers with two or more jobs are properly selected. The inflated number of observations in 85 disappear in 86, so they are incorrectly counted as separations. In the DDD methodology, adopting separations from 85 to 86 as the *before* counterfactual to the treated separations of 86 to 87 would make treatment artificially shrink in comparison (because separations of 1985 are wrongly estimated), leading to an underestimation of the MW effects.[25] Figure 9 shows the magnitude of this issue. On the left-hand side, we see the absolute change in observations before and after sample restrictions are applied and on the right-hand side its variation rate.

[25] If instead of the algorithm, the econometrician resorts to the complete duplicate elimination, this problem is exacerbated.

Another problem is connecting workers from 1985 to 1986. Because the ID is not consistent, one must use firm ID and other worker's characteristics to connect the workers. However, I attempted this method but only managed to connect a smaller percentage of workers, and lose all the workers that change firm in 1985.

Figure 9: Data Issue of 1985



Notes: This figure shows the data cleaning of Quadros de Pessoal for 1985. The panel on the left shows the raw number of observations, before and after cleaning. The panel on the right shows the variation in percentages of the data from raw to clean. Because 1985 misses the cleaning of the duplicates by worker ID, has an inflated number of observations compared to the following years. All years have cleaning based on absurd values of tenure, age, further restrictions to full-time workers, dependent-employed workers, workers in the primary sector, workers with a job in Madeira or Azores.

Can this sampling issue be corrected?

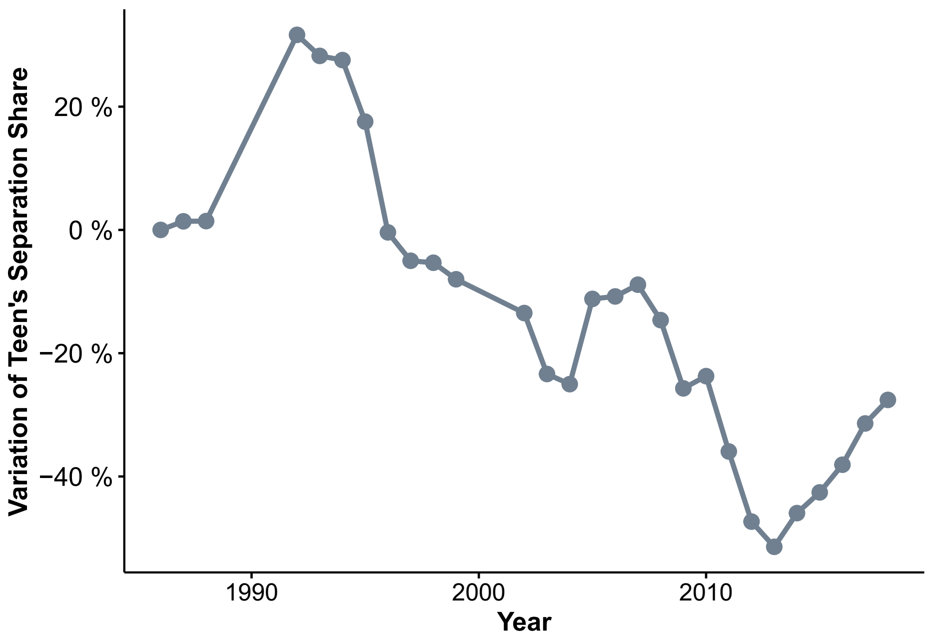
The problem of *overcounting* workers can't be overcome with Dubin and Rivers (1989) sampling bias correction (which is the logit model equivalent of the Heckman (1979) model). Firstly, there isn't a pattern of duplicates to be discovered, since there is only one problematic year. Secondly, because the issue is not about "data missingness", I would need to adapt Dubin and Rivers (1989) model.

A.3 Empirical Application and Strategy

A.3.1 Teen Trends

This Annex argues that there are teen-specific trends that, if otherwise not accounted, will bias this thesis's analysis. To do so, I *semi-replicate* and extend the time-frame of a Poisson regression done by Portugal and Cardoso (2006). Figure 10 plots the change, compared with 1986, of the share of teens (18 and 19 years old) in all separations within young workers from 18 to 30 years old. The share is controlled for firm size, market concentration and sector. For example, in 2011 the share of teens in total young workers separations was half of what it was in 86.

Figure 10: Teen Share of Young Workers' Separations



Notes: This Figure shows the percentual change, compared with 1986, of the share of teens in all separations within young workers from 18 to 30 years old. Teens are considered between 18 and 19 years old. The coefficients are from the log Poisson regression 41, which is controlling for firm size, market concentration and sector. The Herfindahl index is used to measure market concentration and firm size is in the natural log. The sample excludes workers from Madeira or Azores, in the primary sector, with a part-time job and independently employed. Source: Quadros de Pessoal.

$$\frac{teen\ separations_{jt}}{young\ separations_{jt}} = exp\{\vartheta_0 + \sum_{t=1987}^{2019}[\vartheta_{1t}t] + \vartheta_2 log(firmsize_{jt}) + \vartheta_3 H_{S(j)t} + \phi_{S(j)} + u_{jt}\}$$

$$\Leftrightarrow log(teen\ separations_{jt}) = \vartheta_0 + \sum_{t=1987}^{2019}[\vartheta_{1t}t] + \vartheta_2 log(firm\ size_{jt}) + \vartheta_3 H_{S(j)t} + \phi_{S(j)} + log(young\ separations_{jt}) + u_{jt}, \quad (41)$$

where t are the years j are firms, S are the sectors, $firmsize_J$ is the number of employees, $H_{S(j)}$ is the Herfindahl index and $\phi_{S(j)}$ are FE of sectors estimated as dummies. $log(youngseparations_{Jt})$ is the exposure variable.

The estimates in Figure 10 are the time dummies ϑ_{1t} from equation (41), which is a Poisson regression estimated with maximum likelihood. It's clear that teens share in separations is not constant.

A.3.2 Robustness Checks

Table 8: Robustness and Falsification Tests for the Uncontrolled Specification

Nº	Description of the test	DDD α estimate (stand. error)	p-value	95% confidence interval	
				lower bound	upper bound
	Base Estimates	**0.2077** (0.0483)	< 0.001	0.1130	0.3024
	Robustness Checks				
1	Besides MW workers, include workers below MW in the treatment group	**0.2176** (0.0376)	< 0.001	0.1461	0.2892
2	Control age bin changed to age 20-24, instead of 25-29	**0.1512** (0.0466)	0.003	0.0597	0.2426
3	Dependent variable changed to "worker leaves QP in the following year"	**0.1413** (0.0510)	0.005	0.0431	0.2395
4	Control year changed to 1991, instead of 1988	**0.2053** (0.0454)	< 0.001	0.1163	0.2944
5	Bandwidth changed to 2€ each side	**0.3774** (0.0729)	< 0.001	0.2345	0.5204
	Falsification Tests				
6	Treatment age changed to 25-29 and the control to 30-34	0.0539 (0.0288)	**0.0617**	-0.0026	0.1105
7	Treatment wage bin changed to "between the MW and below the 40 percentile"	0.0390 (0.0489)	**0.4250**	-0.0569	0.1349

All p-values and confidence intervals are calculated with cluster robust standard errors at the worker and bins level (more details in Section 5).

This section checks the robustness of the estimates of the minimum wage effect on displacement. Table 8 reports estimates from the key parameter $\alpha_8 8$ of the DDD model (27) in a number of robustness and falsification tests:[26]

1. The most relevant one. It verifies whether the imperfect bindingness of the natural experiment was selective, i.e., employers raised to the MW less separation-prone workers and left workers with higher separation proclivity below the MW, breaking *Assumption 1*. The α coefficient barely changes and it's not statistically different from the base parameter, rejecting the latter hypothesis;
2. To check the robustness of the age counterfactual. The α coefficient is lower but not statistically different from the base parameter;[27]
3. To check the differences between having "worker leaves QP in the following year" instead of separations as the dependent variable. The α coefficient is lower but not statistically different from the base parameter;
4. Meant to further test *Assumption 1*. The α coefficient barely changes and it's not statistically different from the base parameter;
5. Meant to further test the big bandwidth size is an issue. Normally, the bandwidth is smaller. I extended it to 5€ because in 1987 many workers were "around"/below the MW, and I wanted to keep them in the analysis. The α coefficient is lower but not statistically different from the base parameter;
6. To test whether a causal effect between age groups not affected by the treatment is detected by this research design. The α coefficient is not statistically different from 0.
7. Meant to test whether a causal effect between wage groups not directly affected by the treatment is detected by this research design. The α coefficient is not statistically different from 0.

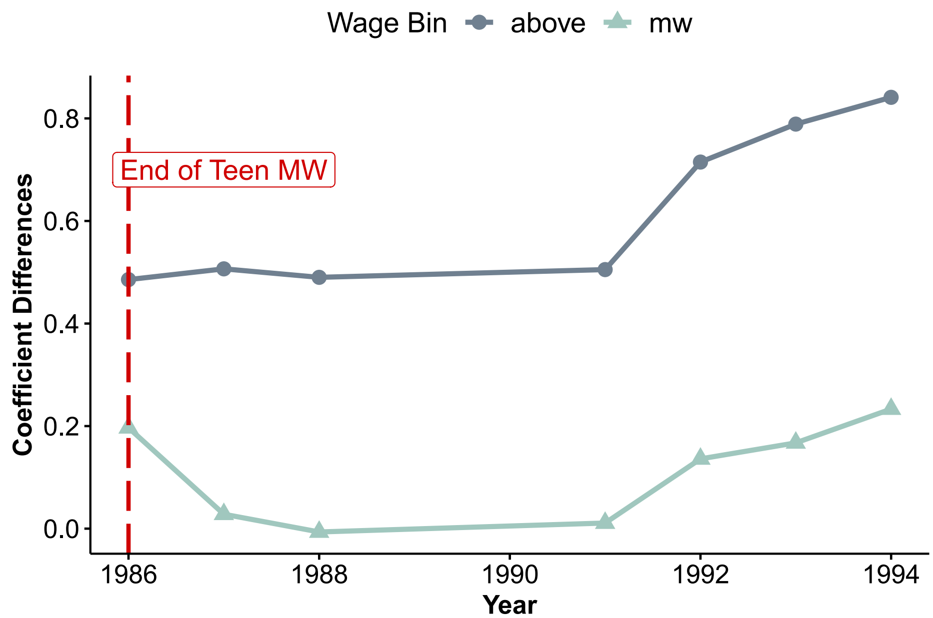
A.3.3 Verifying Assumption 2: Parallel Trends

From Figures 11 and 5 we see that parallel trends hold after the policy is implemented, both in the base and the full model. A well-specified DDD/DiD maintains the *after-treatment* parallel trend both in the controlled and uncontrolled models. When only the magnitude of their differences changes and parallel trends hold both in the base and full model it implies robustness of the counterfactual group. It means that is good enough to encompass most changes by itself (see more reasons in Roth et al. (2022)).

[26] I use the unconditional benchmark because the controlled one is easier to falsely pass a falsification test since the fixed effects "smoothen" the results.

[27] Leaving no confirmation in favor of Pereira (2003) and her findings of age spillovers.

Figure 11: Triple Difference of Separation propensity: Base Values



Notes: This figure reports estimates from the full model 27. The differences between the lines are the α estimates. Every dot is computed by: the logit coefficient of teens minus the logit coefficient of young workers, within each wage bin. Thus, a negative value indicates that young workers of that wage bin are more likely to separate than teen workers. The wage bins refer to workers at the MW and above the 40th percentile of wages of workers above the MW. In 1986, the MW for teen workers was 75% lower, so the separations in the red line, which are separations from 86 to 87, show a MW shock to young workers. Control variables: tenure, firm size, worker and firm fixed effects. The MW workers are defined from a bandwidth of 5€ for each side of the nominal MW. The sample excludes workers from Madeira or Azores, in the primary sector, with a part-time job and independently employed. Source: Quadros de Pessoal.

A.3.4 Estimation

To estimate (26) I compute dummies for all combinations of the dummies (a variable D, which also contains year fixed effects as dummies) because it's less computationally heavy than computing interaction terms. It is also easier to compute the Average Partial Effects for each coefficient. This modification does not change the results in any way, since the procedures are equivalent (Olden and Møen 2022). The α from model (26) is computed as follows: first, I take the age differences (teen minus young workers), then the differences of wage bins (MW minus above), and finally the difference from the years (year t minus 1986). Second, not all workers in QP are included in the sample. Besides perfectly classified and singletons being eliminated, only workers that at any given moment pass by a bin relevant to the analysis keep their ID, while the other workers get a common worker ID (this is done for computational reasons). $579\,235$ different workers pass through some of the interest groups. All the others have a common worker ID. The FE are estimated until 2019 to diminish IPP by catching the maximum number of years possible for each worker.

I use the logit model because, in antinomy with Linear probability models (LPM), it allows for the MW impact on the probability of separation to vary from worker to worker.[28] I.e., the 0.39 logit log-odds effect of the MW will translate into different impacts on the probability of separation $Pr(separations_{it} = 1|\boldsymbol{X})$ (hence the common usage of APE for interpretation), depending on the other covariates ($log(firm\ size_{J(it)t})$, $tenure_{it}$, firm $\psi_{J(it)}$ and worker FE θ_i), because the relationship between the probability of separation and the log-odds is a standard logistic density function.

Fixed Effects

Model (26) has worker and firm FE to mirror an AKM (Abowd, Kramarz, and Margolis 1999) to the separation paradigm. FEs will capture all firm and worker characteristics that are constant over time, even if not observed.[29] With the FEs inclusion, I must delete all singleton observations and all perfectly classified observations (workers/firms with only 1s or 0s in the dependent variable y. Singletons are removed because it's impossible to retrieve meaningful information about the average behavior of a worker from one observation, and I justify why perfectly classified observations are removed in the next sub-section A.3.5. Model (26) can theoretically be estimated using dummies for each firm and worker, but it's not computationally feasible. Thus, I use Stammann (2018) for the estimation of model (26).

A.3.5 Bias Corrections

Incidental Parameter Problem

Parameters estimated with a Maximum Likelihood Estimator will have a bias discovered by Neyman and Scott (1948) (OLS included). In a simple cross-section, this bias can be asymptotically approximated, using the Taylor series, to IPP bias $= \frac{\text{number parameters}}{\text{sample size}}$ (deduction in Iván Fernández-Val and Weidner (2018)[30]). Normal parameters are *consistent* in the light of this estimator. But if it's an incidental parameter, which FE are, (parameters whose dedicated sample does not increase with a sample increase) it's inconsistent. For example, if a worker FE is included, increasing the number of workers in a sample does not decrease the number of observations captured by each worker FE parameter. Only when increasing the number of years worker FE are consistent (see

[28]Moreover, LPM are also inappropriate due to heteroskedasticity and, most importantly, due to bad fit, creating meaningless estimates that exceed the 0 to 1 bound of probabilities (Wooldridge 2010).

[29]This is the main reason why education is not inserted into the equation. Although QP provides this information, it has some mistakes that need to be corrected using max mode, as recommended by BPlim (Bank of Portugal Microdata Research Laboratory). This makes education constant by worker, and wiped out by worker FE (Wooldridge 2010).

[30]Iván Fernández-Val and Weidner (2018) does a mash-up of all the deductions up to 3 FE and even unbalanced, making the IPP bias formula far more complex than the latter. On the Fixed T vs. non-fixed T deductions, Iván Fernández-Val and Weidner (2018) recommends the econometrician to follow the methods more easily available for each, because they are quite similar. Thus, although the paradigm of Phillips and Moon (1999) and Botosaru and Muris (2017) to unbalanced panels are more appropriate, I stick to Ivan Fernández-Val and Weidner (2016)

Lancaster (2000) for a good review of IPP in panel data). So, if the number of years is not high enough, worker FE is biased and will contaminate all structural parameters.

Think of the IPP from the perspective of the Law of Large numbers: worker FE, just like structural parameters, need a sufficiently high sample to correctly perform statistical inference. In the case of Abowd, Kramarz, and Margolis (1999), which I adapt to logit, we need a high number of years so that worker FE $theta_i$ and firm FE $\psi_{F(i,t)}$ have a large enough sample (in the case of firm FE, a higher firm size also helps its estimation). The biggest problem with the IPP relies on workers. In logit, the problem is greatly exacerbated. That's why the perfectly classified are eliminated, because of the IPP: all 1's would make the FE $+\infty$ and all 0's $-\infty$.

To correct the IPP issues, I use the analytical bias correction derived by Ivan Fernández-Val and Weidner (2016) but employ it with the method of alternating projections of Czarnowske and Stammann (2019). For the bias correction to be valid, a certain average number of observations by bin must be in place. The average worker appearances *per* bin are between 14 and 17, as can be seen in Table 3, which is sufficient by the simulations of Czarnowske and Stammann (2019).[31] Finally, I delete observations of workers and firms that show feedback in the FE. Because some workers and firms have high levels of collinearity their FE become ridiculously high and cancel each other, from a feedback effect (Czarnowske and Stammann 2019). It's an easy situation to detect because the FE gets huge negative values on one of the FEs and huge positive values on the other, -50 on worker FE and 50 on firm FE for example. Less than 150 observations relevant to the DDD are eliminated. After I delete them I re-estimate the models, this time free from this collinearity issue.

Rare event bias correction

Separations are consistently below a 50% rate (see Table 3), and logit models perform worse whilst facing this *Rare Event Problem*.[32] To tackle this issue I employ the posterior rare event bias correction developed by King and Zeng (2001). I use equation (42) to find the bias that I later subtract from the $\hat{\boldsymbol{\beta}}$s. Because of the FE, equation (42) is computationally heavy, so I use Gaure (2013) method to estimate a high-dimensional linear weighted least-squares with worker and firm FE, with ς as the dependent variable.

$$bias(\hat{\boldsymbol{\beta}}) = (X^T W X)^{-1} X^T W \varsigma, \tag{42}$$

where $\hat{\boldsymbol{\beta}}$ are the IPP bias-corrected TWFE logit structural parameters; $\varsigma = 0.5\hat{p}_{it} Q_{vv}$; Q_{vv} are the

[31] The quality of bias correction will not only be influenced by the average number of observations, but also by the specific pattern of the dataset.

[32] Contrary to the name, the issue isn't in the rarity, but rather in the possibility of a small absolute number of events. A thousand observations with 10 events is bad, but a million with 10000 events is great. Therefore, the larger problem relies in the estimation of the FEs which then contaminate the bins' coefficients

diagonal elements of $Q = X(X^TWX)^{-1}X^T$; W is a diagonal matrix of $\hat{p}_{it}(1-\hat{p}_{it})$; and $\hat{p}_{it}$ the predicted probabilities.

A.3.6 Average Partial Effects

Log-odds have a difficult interpretation besides the "propensity to separate". The level of all the control variables and FEs of model (26) will impact the probability entailed by the logit log-odds because of the non-linearity of logit. I use APEs, the average probability effect of a given variable on this specific sample.[33] Also, in logit models, the distribution of the residuals has a rescaling effect on parameters (Mood (2009), section 2 explains it). Then, the discrete variable *bins* D may absorb different parts of the residual per level, having a different rescaling and compromising group comparisons between the bins. A literature of complex methods to allow group comparisons exists (see Allison (1999) and Williams (2009)), but the most useful and simplest method is found in Wooldridge (2010): use APEs. The differences in rescaling get diluted, becoming close to irrelevant. Moreover, the FEs also absorb differences in the unobserved constant firm and worker components in the residuals, diminishing the differences in rescaling (Halaby 2004). And for the covariance matrix of the APEs I use the finite sample estimator by Cruz-Gonzalez, Fernández-Val, and Weidner (2017).

Partial Effects on Separations

The *true* probability of separation from model (26) is:

$$\begin{aligned} p_{it} =& Pr(\sum_{t=1987}^{2019}[\beta_{1t}t + \beta_{2t}teen_{it} + \beta_{3t}mw_{it} + \\ & \beta_{4t}(t \times teen_{it}) + \beta_{5t}(t \times mw_{it}) + \beta_{6t}(teen_{it} \times mw_{it}) + \\ & \alpha_t(t \times teen_{it} \times mw_{it})] + \\ & \xi_1 ln(firm\ size_{J(it)t}) + \xi_2 tenure_{it} + \theta_i + \psi_{J(it)} \leq u_{it}) \end{aligned} \tag{43}$$

Because u_{it} is coerced into a standard logit distribution $\varepsilon_{it} \sim logitic(0, \sigma^2_\varepsilon)$:

$$\begin{aligned} Pr(y_{it} = 1|\boldsymbol{X}) =& F(\sum_{t=1987}^{2019}[\beta_{1t}t + \beta_{2t}teen_{it} + \beta_{3t}mw_{it} + \\ & \beta_{1t}(t \times teen_{it}) + \beta_{2t}(t \times mw_{it}) + \beta_{3t}(teen_{it} \times mw_{it}) + \\ & \alpha_t(t \times teen_{it} \times mw_{it})] + \\ & \xi_1 ln(firm\ size_{J(it)t}) + \xi_2 tenure_{it} + \theta_i + \psi_{J(it)}) \end{aligned} \tag{44}$$

[33] Some authors call this Average Medium Effects (Mood 2009). I'm using Wooldridge (2010) naming.

Where F is the standard logit cumulative probability function and $Pr()$ the estimated probability. Thus, the partial effects of the continuous variables:

$$\frac{\partial Pr(y_{it} = 1|\boldsymbol{X})}{\partial firm\ size_{it}} = \frac{\partial F(X\beta)}{\partial firm\ size_{it}} = \frac{\partial F(X\beta)}{\partial XB}\frac{\partial X\beta}{\partial firm\ size_{it}} = f(X\beta)\xi_1$$

$$\frac{\partial Pr(y_{it} = 1|\boldsymbol{X})}{\partial tenure_{it}} = \frac{\partial F(X\beta)}{\partial tenure_{it}} = \frac{\partial F(X\beta)}{\partial XB}\frac{\partial X\beta}{\partial tenure_{it}} = f(X\beta)\xi_2$$

where $f()$ is the standard logit density distribution and $\boldsymbol{X\beta}$ represents all variables. And for the discrete variables, exemplified for α_{88} (the DDD coefficient of interest):

$$F(X\beta) - F(X\beta - \alpha_{88})$$

From the partial effects, we can apply an average and compute the average partial effects. Table 4 shows the APE's of α_{88}, $firm\ size_{it}$ and $tenure_{it}$:

$$APE_{MW} = \frac{\sum F(X\beta) - F(X\beta - \alpha_{88})}{N_1}$$

$$APE_{firm\ size} = \frac{\sum f(X\beta)\xi_1}{N}$$

$$APE_{tenure} = \frac{\sum f(X\beta)\xi_2}{N}$$

where N is the number of observations and $N1$ is the number of observations used to estimate α_{88}.

A.3.7 Full Coefficients Results

This section shows all the coefficients from the main regression of the empirical application.

Table 9: Full Models estimated with the D variable.

Year	Wage	Age	Full Model without Bias Corrections	Bias-Corrected Full Model
1991	MW	25-29	-0.05^{*}	-0.06^{*}
			$[-0.08; -0.02]$	$[-0.10; -0.03]$
1986	MW	25-29	-0.85^{*}	-0.75^{*}
			$[-0.88; -0.81]$	$[-0.79; -0.72]$
1992	Above	25-29	0.18^{*}	0.15^{*}
			$[0.17; 0.20]$	$[0.13; 0.16]$
1991	Above	25-29	-0.07^{*}	-0.08^{*}
			$[-0.09; -0.06]$	$[-0.10; -0.07]$
1987	Above	25-29	-0.16^{*}	-0.17^{*}
			$[-0.18; -0.14]$	$[-0.19; -0.15]$
1988	MW	25-29	0.19^{*}	0.15^{*}
			$[0.16; 0.23]$	$[0.11; 0.18]$
1994	Above	25-29	-0.09^{*}	-0.09^{*}
			$[-0.10; -0.07]$	$[-0.11; -0.08]$
1988	Above	25-29	0.03^{*}	0.01
			$[0.01; 0.05]$	$[-0.01; 0.02]$
1986	Above	25-29	-0.82^{*}	-0.74^{*}
			$[-0.84; -0.80]$	$[-0.76; -0.72]$
1993	MW	25-29	0.65^{*}	0.57^{*}
			$[0.62; 0.69]$	$[0.53; 0.61]$
1987	MW	25-29	-0.07^{*}	-0.09^{*}
			$[-0.10; -0.03]$	$[-0.12; -0.05]$
1993	Above	25-29	0.63^{*}	0.55^{*}
			$[0.62; 0.65]$	$[0.54; 0.57]$
1992	MW	18-19	-0.01	-0.03
			$[-0.07; 0.05]$	$[-0.09; 0.02]$
1991	Above	18-19	-0.09^{*}	-0.10^{*}
			$[-0.14; -0.04]$	$[-0.15; -0.05]$
1988	Above	18-19	-0.02	-0.05
			$[-0.09; 0.04]$	$[-0.11; 0.02]$
1994	MW	25-29	-0.01	-0.03
			$[-0.05; 0.02]$	$[-0.06; 0.01]$
1991	MW	18-19	-0.36^{*}	-0.34^{*}
			$[-0.41; -0.30]$	$[-0.40; -0.29]$
1987	MW	18-19	-0.41^{*}	-0.39^{*}
			$[-0.47; -0.35]$	$[-0.45; -0.32]$
1994	MW	18-19	-0.03	-0.05
			$[-0.09; 0.03]$	$[-0.11; 0.01]$
1987	Above	18-19	-0.20^{*}	-0.21^{*}
			$[-0.27; -0.13]$	$[-0.28; -0.14]$
1986	MW	18-19	-0.88^{*}	-0.79^{*}
			$[-0.95; -0.81]$	$[-0.86; -0.72]$
1994	Above	18-19	0.34^{*}	0.27^{*}
			$[0.29; 0.38]$	$[0.23; 0.32]$
1993	MW	18-19	0.51^{*}	0.43^{*}
			$[0.45; 0.58]$	$[0.37; 0.50]$
1992	MW	25-29	0.22^{*}	0.18^{*}
			$[0.18; 0.25]$	$[0.14; 0.21]$
1988	MW	18-19	-0.19^{*}	-0.19^{*}
			$[-0.25; -0.12]$	$[-0.26; -0.12]$
1992	Above	18-19	0.38^{*}	0.32^{*}
			$[0.34; 0.43]$	$[0.27; 0.36]$
1993	Above	18-19	0.98^{*}	0.84^{*}
			$[0.93; 1.03]$	$[0.79; 0.89]$
1986	Above	18-19	-1.00^{*}	-0.88^{*}
			$[-1.05; -0.95]$	$[-0.93; -0.83]$
log(firm size)			-0.02^{*}	-0.02^{*}
			$[-0.02; -0.02]$	$[-0.02; -0.02]$
tenure			-0.00^{*}	-0.00^{*}
			$[-0.00; -0.00]$	$[-0.00; -0.00]$
Deviance			61593462.09	61593462.09
Num. obs.			59260884	59260884
Num. groups: worker			579236	579236
Num. groups: firm			678054	678054

* Null hypothesis value outside the confidence interval of 95%. Standard errors are calculated with clusters on workers, firms and bins. With bias corrections.

Table 10: APE from bias-corrected full model.

Year	Wage	Age	APE
1991	MW	25-29	-0.01094e-02***
1986	MW	25-29	$-1.107e-01^{***}$
1992	Above	25-29	2.625e-02***
1991	Above	25-29	-1.396e-02***
1987	Above	25-29	-2.813e-02***
1988	MW	25-29	2.571e-02***
1994	Above	25-29	-1.570e-02***
1988	Above	25-29	1.087e-03
1986	Above	25-29	-1.091e-01***
1993	MW	25-29	1.073e-01***
1987	MW	25-29	-1.468e-02***
1993	MW	25-29	1.046e-01***
1992	MW	18-19	-5.893e-03
1991	Above	18-19	-1.713e-02***
1988	Above	18-19	-8.103e-03
1994	MW	25-29	-4.414e-03
1991	MW	18-19	-5.546e-02***
1987	MW	18-19	-6.189e-02***
1994	MW	18-19	-8.550e-03
1987	Above	18-19	-3.491e-02***
1986	MW	18-19	-1.148e-01***
1994	Above	18-19	4.965e-02 ***
1993	MW	18-19	7.978e-02***
1992	MW	25-29	3.118e-02 ***
1988	MW	18-19	-3.171e-02 ***
1992	Above	18-19	5.784e-02 ***
1993	Above	18-19	1.643e-01 ***
1986	Above	18-19	-1.251e-01 ***
log(firm size)			-4.010e-03 ***
tenure			-3.864e-04 ***

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors are in parenthesis and calculated with clusters on workers, firms and bins. With bias corrections.

A.3.8 Demonstration Linear Gelbach

This annex proves the Gelbach decomposition in a linear setting, the one that is extrapolated to the logit model and used in table 6. Equation (45) is similar to equation (26) but with separations y_{it} directly as a dependent variable (and the change for the $\boldsymbol{D}$ variable.

$$\mathbf{Y} = \mathbf{D}\boldsymbol{\beta}^{LPM} + \textbf{firm size}\,\boldsymbol{\xi}_1^{LPM} + \textbf{tenure}\,\boldsymbol{\xi}_2^{LPM} + \mathbf{M}_1\theta^{LPM} + \mathbf{M}_2\psi^{LPM} + \mathbf{u}^{LPM} \tag{45}$$

Because linear methods and will estimate $\hat{\boldsymbol{\beta}} E\left[[\boldsymbol{D'D}]^{-1}\boldsymbol{D'Y}\right]$, by multiplying the estimated (45) by $(D'D)^{-1}D'$ on both sides we reach the decomposition:

$$(\boldsymbol{D'D})^{-1}\boldsymbol{D'Y} = (\boldsymbol{D'D})^{-1}\boldsymbol{D'D}\hat{\boldsymbol{\beta}} + (\boldsymbol{D'D})^{-1}\boldsymbol{D'}\,\boldsymbol{firm\ size}\,\hat{\xi}_1 + (\boldsymbol{D'D})^{-1}\boldsymbol{D'}\,\mathbf{tenure}\,\hat{\xi}_2 + \\ (\boldsymbol{D'D})^{-1}\boldsymbol{D'M}_1\hat{\theta} + (D'D)^{-1}D'\boldsymbol{M}_2\hat{\psi} \quad (46)$$

Considering linear adaptations of the base model (27) and the auxiliary regressions (28) (29) (30) (31):

$$\hat{\beta}_{base} = \hat{\beta}_{full} + \hat{\Gamma}^{tenure}\hat{\xi}_1 + \hat{\Gamma}^{firm\ size}\hat{\xi}_2 + \hat{\Gamma}^{\theta}M_1\hat{\theta} + \hat{\Gamma}^{\psi}M_1\hat{\psi}$$
$$bi\hat{a}s\ confounding = \hat{\Gamma}^{tenure}\hat{\xi}_1 + \hat{\Gamma}^{firm\ size}\hat{\xi}_2 + \hat{\Gamma}^{\theta}M_1\hat{\theta} + \hat{\Gamma}^{\psi}M_1\hat{\psi}$$

A.3.9 Estimation of the Precision Matrix

There are several methods available to estimate a precision matrix from data following a multi-normal distribution. See Fan, Liao, and Liu (2016) for an excellent review of several methods. However, most methods yield very similar results. Both `R`, `STATA` and `Python` have packages that estimate it efficiently.

I´ll use the method developed by Yuan and Lin (2007). I further estimated the precision matrix for the application with the methods developed by Banerjee et al. (2006) and K. Lee and Lee (2017), and the results were almost identical.

A.3.10 Diagnostics

The rigor of the decomposition of the rescaling effects depends on the normality of the controls. This section provides a visual representation of some residuals used to do the CAR decomposition in section 4.6, the ones from 1986. Although all the controls fail normality tests, they may be sufficiently close.

Worker FE is figure 12 and firm FE is figure 13. Both are represented in a density plot. The red line represents a normal distribution with the same mean and standard deviation as the sample. We can see that there is a clear deviation from the normal distribution. Tenure and log of firm size (figure 14 and 15, respectively) are better represented in qqplots, where it's clear the distributions are closer to a log-normal distribution.

Figure 12: Density plots of 1986 Worker FE residuals from Equation 26

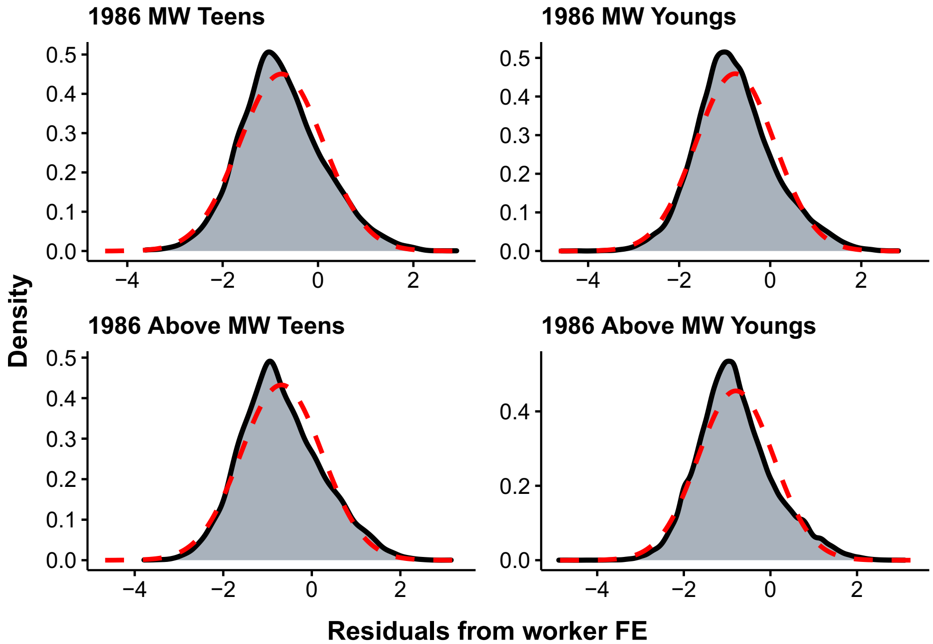


Figure 13: Density plots of 1986 Firm FE residuals from Equation 26

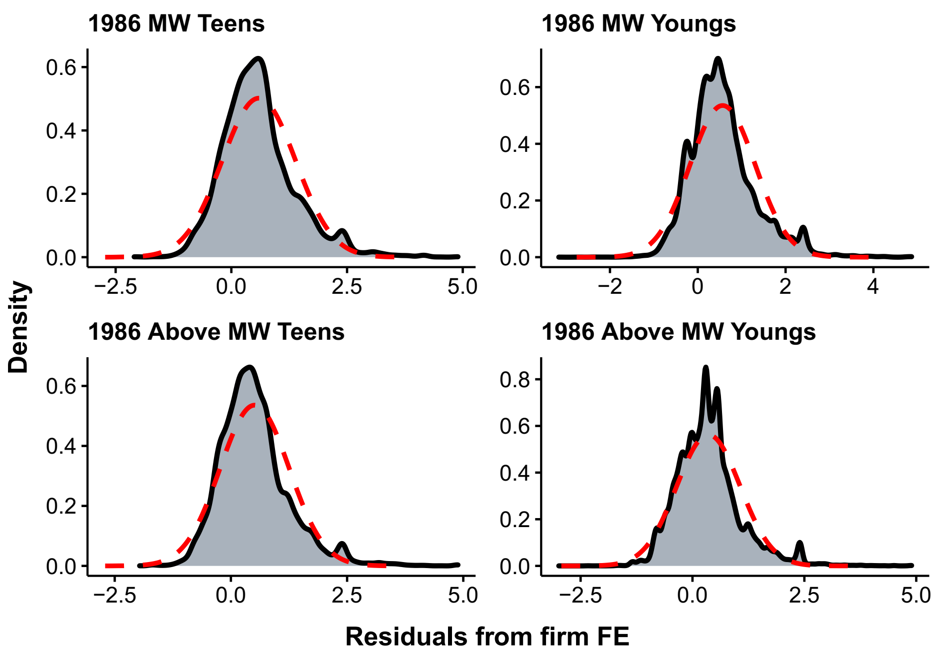


Figure 14: Density plots of 1986 Firm FE residuals from Equation 26

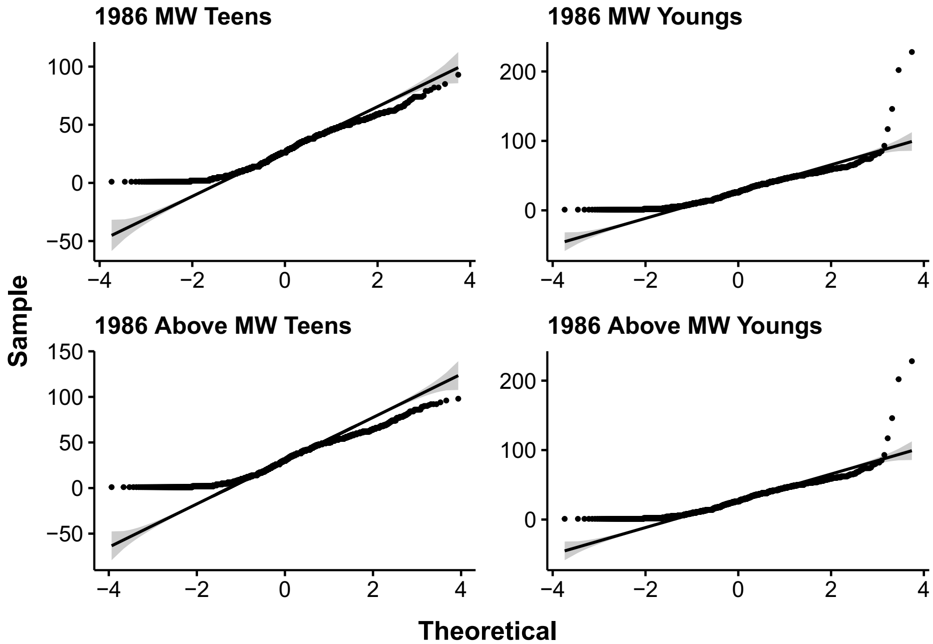
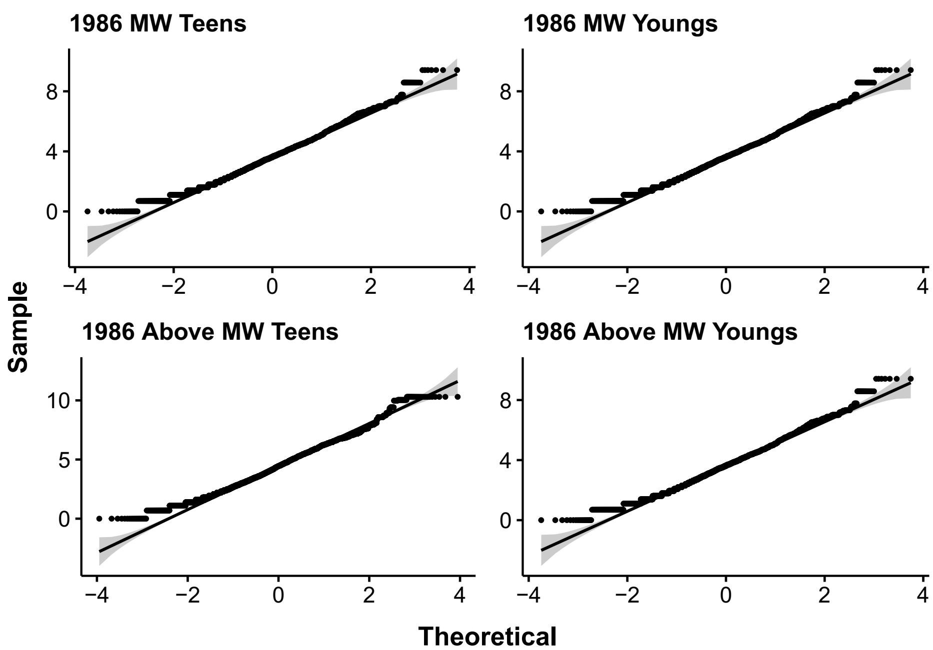


Figure 15: Density plots of 1986 ln(firm size) residuals from Equation 26



A.4 Portuguese Institutional Setting

Countries may have institutions that interact with MW policy changing its effect.[34] Therefore, it's plausible that the Portuguese labour market is more monopsonistic than others. The strong employment protection, the wage rigidity, and the wage setting system interact with MW and alter internal labour markets.

The biggest labour institution in Portugal is the collective bargaining system. In Portugal, massive collective agreements are published in *Boletim de Trabalho e Emprego* and, if they fulfill certain criteria about the representation of workers of a sector, the State may extend collective agreements to all workers of that sector. As a result, 90% of the private sector is covered under these agreements (Addison, Portugal, and Vilares 2017), even in firms that don't bargain and to non-unionized workers. In sum, the State creates other minimum wages; more specifically, around 30000 minimum wages (Martins 2014).

A.4.1 Was the MW binding?

Table 11: Exceptions to the MW law of 1987

Designation	Condition for the Exception	Amount of the MW
Apprentice	Workers below 25 years old A situation characterized by some level of on-the-job training Cannot exceed 2 years The worker cannot have a technical diploma in that area	80%
Agriculture	CAE 1 sector	0.89%
Domestic	Domestic/cleaning services provided by non-specialized firms	0.69%
Small Firms	Firms with 5 or fewer workers	0.89%
Craftsmen	Activities of artisanal nature	Different legal status
Medium Firms	Firms with 6 to 50 workers and Wage expenditure increase superior to 80% of the MW update	Granting pending on MTSSS acceptance.

Source: Decreto-Lei n.º 69-A/87, Diário da Republica. The agricultural exception does not affect my results because the entire sector is deleted from the sample, due to its unreliability. MTSSS is the Ministry of Labour, Social Security and Solidarity. The government had 90 days to notify the medium firms that asked for the exception if it was granted, and that time exceeds the QP survey filling (March of 1987).

Given the previous information, I take caution to identify the bindingness of the minimum wage (MW). In this section, I show that the MW in fact binding for MW teen workers. Portugal and

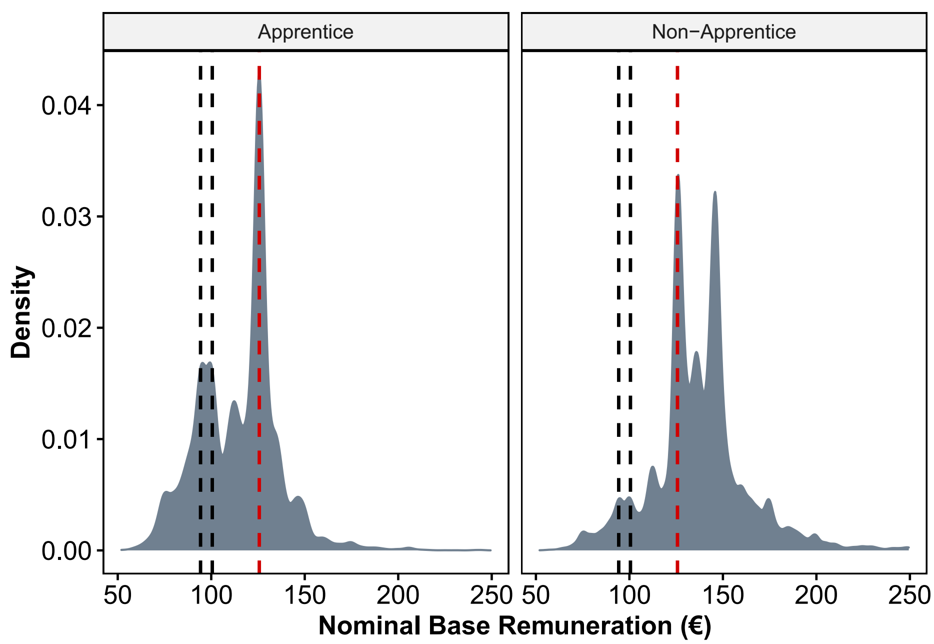
[34] Neumark and Wascher (2004) compare OECD countries showing some institutions (as employment protection, union coverage and active labour market policies) interact with MW policies; Boockmann (2010) shows that employment protection, union coverage and active labor market policies explain part of these differences

Cardoso (2006) had already concluded that the natural experiment we explore is binding for MW workers.

In Figure 4 is possible to see a big jump in teen MW workers following the new MW. However, the lump to the left of the MW still raises some questions. It's the result of several exceptions engraved in the MW law *Decreto-Lei n.º 69-A/87, Diário da Republica*. I summarize those exceptions in Table 11. Thus, the *lump* is excepted and not a weakness of this experiment.

The apprentice exception deserves some special attention. Could some employers have shifted the classification of their employees to apprentice to avoid the MW hike? And, to avoid jeopardizing the whole experiment, why don't I delete them (in QP it's possible to know if a worker was an apprentice or not)? Firstly, apprentices are 55.6% of all 18 and 19 years old in 1986. Secondly, as shown in Figure 16, the distribution of apprentices and non-apprentices are quite similar. They only differ in high-wage jobs. This means that most apprentices earn the MW and not the apprentice MW (which is 80%.

Figure 16: Wages in 1987 for Teens by Apprenticeship



Notes: This figure shows the impact of the minimum wage (MW) on the wage distribution of teens by apprenticeship. Base wages, which exclude all benefits, overtime payments, and indemnifications, are the ones bound by the MW. The sample excludes workers from Madeira or Azores, in the primary sector, with a part-time job and independently employed. The red dashed lines are the minimum wages. The black dashed lines are 75% and 80% of MW, respectively from the left hand side to the right. Sources: Quadros de Pessoal for wage data; Instituto Nacional de Estatística, for the minimum wage data.

A.5 Further Proofs for the Logit Decomposition

A.5.1 1[st]

$$\frac{P(\hat{\boldsymbol{\upsilon}}|x,y=0)}{P(\hat{\boldsymbol{\upsilon}}|x,y=1)} = e^{\frac{1}{2}\left[(\hat{\boldsymbol{\upsilon}}-\hat{\boldsymbol{\delta}}_0-\hat{\boldsymbol{\delta}}_1\mathbf{X}-\hat{\boldsymbol{\delta}}_2)^{\intercal}\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\upsilon}}-\hat{\boldsymbol{\delta}}_0-\hat{\boldsymbol{\delta}}_1\mathbf{X}-\hat{\boldsymbol{\delta}}_2)-(\hat{\boldsymbol{\upsilon}}-\hat{\boldsymbol{\delta}}_0-\hat{\boldsymbol{\delta}}_1\mathbf{X})^{\intercal}\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\upsilon}}-\hat{\boldsymbol{\delta}}_0-\hat{\boldsymbol{\delta}}_1\mathbf{X})\right]}$$

using K to simplify notation: $K^p_{y=1} = \hat{\upsilon}^p - \hat{\delta}^p_0 - \hat{\delta}^p_1 x - \hat{\delta}^p_2$ and $K^p_{y=0} = \hat{\upsilon}^p - \hat{\delta}^p_0 - \hat{\delta}^p_1 x$

$$\frac{1}{2}\left[\begin{array}{cccc} K^1_{y=1} & K^2_{y=1} & ... & K^p_{y=1} \end{array}\right]\left[\begin{array}{cccc} \xi_{11} & - & - & - \\ \xi_{12} & \xi_{22} & - & - \\ ... & ... & ... & - \\ \xi_{1p} & \xi_{2p} & ... & \xi_{pp} \end{array}\right]\left[\begin{array}{c} K^1_{y=1} \\ K^2_{y=1} \\ ... \\ K^p_{y=1} \end{array}\right] - \frac{1}{2}\left[\begin{array}{cccc} K^1_{y=0} & K^2_{y=0} & ... & K^p_{y=0} \end{array}\right]\left[\begin{array}{cccc} \xi_{11} & - & - & - \\ \xi_{12} & \xi_{22} & - & - \\ ... & ... & ... & - \\ \xi_{1p} & \xi_{2p} & ... & \xi_{pp} \end{array}\right]\left[\begin{array}{c} K^1_{y=0} \\ K^2_{y=0} \\ ... \\ K^p_{y=0} \end{array}\right] = \tag{47}$$

$$\frac{1}{2}\left[\begin{array}{cccc} \sum_{z=1}^p(K^z_{y=1}\xi_{1z}) & \sum_{z=1}^p(K^z_{y=1}\xi_{2z}) & ... & \sum_{z=1}^p(K^z_{y=1}\xi_{pz}) \end{array}\right]\left[\begin{array}{c} K^1_{y=1} \\ K^2_{y=1} \\ ... \\ K^p_{y=1} \end{array}\right] - \frac{1}{2}\left[\begin{array}{cccc} \sum_{z=1}^p(K^z_{y=0}\xi_{1z}) & \sum_{z=1}^p(K^z_{y=0}\xi_{2z}) & ... & \sum_{z=1}^p(K^z_{y=0}\xi_{pz}) \end{array}\right]\left[\begin{array}{c} K^1_{y=0} \\ K^2_{y=0} \\ ... \\ K^p_{y=0} \end{array}\right] = \tag{48}$$

$$\frac{1}{2}\sum_{z=1}^{p}\left[K^z_{y=1}\sum_{\dot{z}=1}^{p}(K^{\dot{z}}_{y=1}\xi_{z\dot{z}})\right] - \frac{1}{2}\sum_{z=1}^{p}\left[K^z_{y=0}\sum_{\dot{z}=1}^{p}(K^{\dot{z}}_{y=0}\xi_{z\dot{z}})\right] =$$

where $\dot{z}$ is . In this equation, the only parts that survive are the ones that depend on $\hat{\delta}^z_2 \; \forall \; z \in [1:p]$ appearing on $K_{y=1}$'s. Use the squares property $a^2 - b^2 = (a-b)(a+b)$:

$$\frac{1}{2}\sum_{z=1}^{p}\left[-\widehat{\delta}_2^z\sum_{\dot{z}=1}^{p}((\widehat{\upsilon}^{\dot{z}}-\widehat{\delta}_0^{\dot{z}}-\widehat{\delta}_1^{\dot{z}}x-\widehat{\delta}_2^{\dot{z}})\xi_{z\dot{z}})\right]+\frac{1}{2}\sum_{z=1}^{p}\left[(\widehat{\upsilon}^z-\widehat{\delta}_0^z-\widehat{\delta}_1^z x)\sum_{\dot{z}=1}^{p}(-\widehat{\delta}_2^{\dot{z}}\xi_{z\dot{z}})\right]=$$

Because $\mathbf{\Sigma}^{-1}$ is symmetric and $\dot{z}$ and z are interchangeable: $\sum_{z=1}^{p}\left[(\upsilon^z-\delta_0^z-\delta_1^z x)\sum_{\dot{z}=1}^{p}(-\delta_2^{\dot{z}}\xi_{z\dot{z}})\right]=\sum_{\dot{z}=1}^{p}\left[(\upsilon^{\dot{z}}-\delta_0^{\dot{z}}-\delta_1^{\dot{z}}x)\sum_{z=1}^{p}(-\delta_2^z\xi_{z\dot{z}})\right]$. We can see that the second term is a repetition of the 1st term, except for $-\delta_2^{\dot{z}}$.

$$\sum_{z=1}^{p}\left[\widehat{\delta}_2^z\sum_{\dot{z}=1}^{p}((\widehat{\delta}_0^{\dot{z}}+\widehat{\delta}_1^{\dot{z}}x+\frac{\widehat{\delta}_2^{\dot{z}}}{2}-\widehat{\upsilon}^{\dot{z}})\xi_{z\dot{z}})\right]$$

A.5.2 Independence of $\boldsymbol{\upsilon}$ to base

From equation (18) we can retrieve the $\boldsymbol{\upsilon}$ dependent terms:

$$\sum_{z=1}^{p}\widehat{\upsilon}_i^z\widehat{\theta}^z=\sum_{z=1}^{p}\left[\widehat{\delta}_2^z\sum_{\dot{z}=1}^{p}\widehat{\upsilon}_i^{\dot{z}}\xi_{z\dot{z}}\right]$$

Because $\mathbf{\Sigma}^{-1}$ is symmetric and $\dot{z}$ and z are interchangeable:

$$\Leftrightarrow\sum_{z=1}^{p}\widehat{\upsilon}_i^z\widehat{\theta}^z=\sum_{z=1}^{p}\left[\widehat{\upsilon}_i^z\sum_{\dot{z}=1}^{p}\widehat{\delta}_2^{\dot{z}}\xi_{z\dot{z}}\right]$$

And the need for each variable to individually not influence the base equation directly:

$$\Leftrightarrow 0=\widehat{\theta}^z-\sum_{\dot{z}=1}^{p}\widehat{\delta}_2^{\dot{z}}\xi_{z\dot{z}}\ \vee\ 0=\upsilon_i^z,\quad\forall\, z\in[1:p]$$

disregarding the possibility of all observations of a variable being 0:

$$\Leftrightarrow 0=\widehat{\theta}^z-\widehat{\delta}_2^z\xi_{zz}-\sum_{\dot{z}\neq z}^{p}\widehat{\delta}_2^{\dot{z}}\xi_{z\dot{z}}$$

$$\Leftrightarrow\xi_{zz}=\frac{\widehat{\theta}^z-\sum_{\dot{z}\neq z}^{p}\widehat{\delta}_2^{\dot{z}}\xi_{z\dot{z}}}{\widehat{\delta}_2^z}$$

A.5.3 The "Variance" literature strand

If both full and base models have residuals u that are logistically distributed, then:

Full:

$$\frac{\sqrt{Var(\varepsilon)}}{\sqrt{\hat{V}ar(u^f)}} log\left[\frac{p^f_{it}}{1-p^f_{it}}\right] = \frac{\sqrt{Var(\varepsilon)}}{\sqrt{\hat{V}ar(u^f)}}\beta^f_0 + \frac{\sqrt{Var(\varepsilon)}}{\sqrt{\hat{V}ar(u^f)}}\beta^f_1 x_i + \frac{\sqrt{Var(\varepsilon)}}{\sqrt{\hat{V}ar(u^f)}}\boldsymbol{\beta}_2\mathbf{Z} + \frac{\sqrt{Var(\varepsilon)}}{\sqrt{\hat{V}ar(u^f)}}u^f_i$$

Base:

$$\frac{\sqrt{Var(\varepsilon)}}{\sqrt{\hat{V}ar(u^b)}} log\left[\frac{p^b_{it}}{1-p^b_{it}}\right] = \frac{\sqrt{Var(\varepsilon)}}{\sqrt{\hat{V}ar(u^b)}}\beta^b_0 + \frac{\sqrt{Var(\varepsilon)}}{\sqrt{\hat{V}ar(u^b)}}\beta^b_1 x_i + \frac{\sqrt{Var(\varepsilon)}}{\sqrt{\hat{V}ar(u^b)}}u^b_i$$

Making the bias, and applying Gelbach (2009) to an imaginary *non-rescaled* model:

$$\frac{\sqrt{Var(\varepsilon)}}{\sqrt{\hat{V}ar(u^f)}}\left(\hat{\beta}^f_1 + \sum_{z=1}^{p}\left[\hat{\Gamma}^z_1\hat{\beta}^z_2\right]\right)x_i = \frac{\sqrt{Var(\varepsilon)}}{\sqrt{\hat{V}ar(u^b)}}\hat{\beta}^b_1 x_i$$

$$\hat{\beta}^b_1 = \left(\hat{\beta}^f_1 + \underbrace{\sum_{z=1}^{p}\left[\hat{\Gamma}^z_1\hat{\beta}^z_2\right]}_{\substack{\text{Correlated}\\ \text{bias}}}\right)\underbrace{\frac{\sqrt{Var(u^b)}}{\sqrt{\hat{V}ar(u^f)}}}_{\substack{\text{Uncorrelated}\\ \text{unobserved}\\ \text{heterogeneity}}}, \tag{49}$$

A.6 Literature of the Minimum Wage Unemployment Effects

The usage of natural experiments and DiD is so common in the MW literature that it even has its name: the New Minimum Wage Research (NMWR). It comprises several influential papers, like Card and Krueger (1994), Dube, Lester, and Reich (2010) and Cengiz et al. (2019). These papers all use a natural wedge for their difference in differences: because the U.S. States have the autonomy to raise the MW above the federal one, contiguous States with different minimum wages become natural counterfactuals.

Card and Krueger (1994) were the first to use a difference in differences to analyse the impact of the MW on employment. Their methods, as well as their positive employment results, we controversial. For example, Neumark and Wascher (2000) point out some sampling issues and problems at conducting the survey in Card and Krueger (1994) that invalidate the results. Still, the NMWR survived, because the general methodology was an econometric improvement.

Card and Krueger (1995) summarize the early NMWR results, saying the MW had impacts on employment "around 0". At the time, Card and Krueger (1995) went against the consensus solidified by Brown, Gilroy, and Kohen (1982), who surveyed 25 papers of specifications not using natural experiments and found -0.3 employment elasticity for unskilled workers. More recent meta-analysis point out that both sides (NMWR and the traditional specifications) find, in general, negative employment effects (Neumark and Wascher (2006); Neumark and Shirley (2021)). However, the NMWR tends to find smaller negative results. For a further methodological debate, see Neumark, Salas, and Wascher (2014) and Jardim et al. (2022) defense of traditional specifications, and Allegretto et al. (2017) for a defense of the NMWR.

Even economic theory offers nuance about the effect of the MW on employment. Its impact on employment will depend on the competitiveness of the market (can even be positive in the monopolistic model of Stigler (1946)) and on the absorption taken by prices and profits (see MaCurdy (2015) MW general equilibrium model).

The usage of difference in differences to gauge the employment effects of the MW has also been employed outside of the United States. See, for example, Dolado et al. (1996) for Spain and France, Stewart (2002) for the UK, and the aforementioned studies for Portugal.

About the topic of heterogeneity, some literature adds interaction terms. It does not divide heterogeneity, but it manages to get a sense of the more affected groups. Harasztosi and Lindner (2019) run separate regressions by sector, firm size, or age. Cengiz et al. (2019) and Neumark, Schweitzer, and Wascher (2000) explore the effect of the MW throughout the wage distribution. Burkhauser, Couch, and Wittenburg (1996) checks the effect of household income. MaCurdy (2015) sees if the MW affects more consumers, workers, or investors.