ESSAYS ON ECONOMETRIC METHODS

Annalivia Polselli

A thesis submitted for the degree of

Doctor of Philosophy in Economics

Department of Economics

University of Essex

September 2022

Sapientis autem cogitatio non ferme ad investigandum adhibet oculos advocatos.

(M. T. Cicero, Tuscolanae disputationes, 5, 111)

Acknowledgments

I am very grateful to my supervisors Marcus Chambers and Marco Francesconi for their valuable guidance and support during my PhD.

I also would like to express my sincere gratitude to other researchers who have contributed to my professional development with their helpful comments and stimulating discussions. In particular, Ines Buono, Emilia Del Bono, Emma Duchini, Jayant Ganguli, Abhimanyu Gupta, Shin Kanaya, Riccardo Leoncini, Neslihan Sakarya, and Michel Serafinelli.

I would like to thank friends and colleagues for their active presence. In particular, Ashley Burdett, Francesca Caucci, Nisha Chandhar, Camila Comunello, Ilaria Evangelista, Elena Faieta, Aitor Irastorza-Fadrique, Hamid Nejadghorban, Ludovica Orlandi, Veronica Pizziol, Tommaso Sartori, Antonio Scrocco, Sonkurt Sen, Irina Straton, Elena Toppi, and Teresa Vallefuoco. Special thanks to Mariele Macaluso for her constant support as co-author and true friend, and to Omar Mohsen for his useful suggestions and support over these past months.

Last but not least, I am immensely grateful to my parents, my sister and my extended family for their constant support, unconditional affection and infinite patience during this long journey. They have always been an indispensable part of each of my achievements.

This doctoral thesis is funded by SeNSS and the University of Essex.

Contents

List of Tables

List of Figures

Summary

This thesis consists of three chapters on econometric methods. In Chapter 1, I investigate the consequences of the simultaneous presence of small sample size, leveraged data, and heteroskedastic disturbances on the validity of the statistical inference in linear panel data models. I formalise the panel versions of two jackknife-type estimators and propose a new hybrid estimator. I derive their asymptotic distributions and analyse their finite sample properties with Monte Carlo simulations. I find that test statistics obtained with conventional robust standard errors are over-sized, upward biased, and with less power under heteroskedasticity and with good leveraged data and in small samples. In Chapter 2, I develop diagnostic methods for panel data to detect three types of anomalous units. I formalise statistical measures for quantifying the degree of leverage and outlyingness of units, and develop a method to visually detect the type of anomaly and its effect on other units. I use network analysis tools to show the total and bilateral influence. I then apply my method to four cross-country data sets used in published articles. Chapter 3 investigates the effect of gender sectoral segregation on employment contracts (part-time, permanent, remote work, number of weekly working hours) and hourly wages for both men and women. We use propensity score matching, the Kitagawa-Blinder-Oaxaca decomposition and Mincerian wage regressions to analyse the contribution of observable and unobservable factors on labour outcomes. We find that contractual features systematically chosen by a specific gender are more common in sectors dominated by that group and for both genders. Workers employed in female-dominated sectors are on average paid less but most of the gap is explained by the coefficient effect rather than differences in endowments in both gender dominated sectors. Women self-select into low-paid jobs where their skills are valued less, especially in female-dominated sectors.

Introduction

Applied social science researchers often use econometric methods to conduct causal inference exercises on the phenomenon of interest. They aim to arrive at valid statistical inferences and to consistently estimate the parameters reflecting the underlying relationship. For example, when the assumption of homoskedasticity is violated, econometric models that use asymptotic standard errors produce invalid statistical inferences. This requires the use of robust standard errors (Eicker (1967), Huber et al. (1967), White (1980), and Arellano (1987)). However, the presence of data points that exhibit extreme values in the covariates (*good leverage points*) invalidates statistical inference with robust standard errors (Long and Ervin, 2000; Hayes and Cai, 2007; MacKinnon, 2013; Bramati and Croux, 2007; Verardi and Croux, 2009).

On the other hand, the main objective of applied social scientists is to consistently estimate the causal impact of a treatment on an outcome of interest. This involves constructing a measure comparing the outcomes of treated and untreated units. However, it is not possible to identify appropriate counterfactuals with observational data, given the available set of controls. Researchers use different techniques – such as, differences-in-differences (Card, 1992; Card et al., 1994; Card and Krueger, 2000), synthetic control (Abadie and Gardeazabal, 2003; Abadie et al., 2010, 2015), matching methods (Abadie and Imbens, 2006), matching and synthetic control method (MASC) (Kellogg et al., 2021) – to retrieve the causal estimates if certain assumptions are satisfied. For example, the propensity score matching (PSM) methodology addresses the issue of finding a good counterfactual by matching on the propensity score – i.e., on the conditional probability of receiving the treatment (Rosenbaum and Rubin, 1983).

This thesis consists of three chapters on econometric methods. Chapters 1 and 2 are technical econometric studies on statistical inference, and Chapter 3 applies econometric techniques to an economic issue. Chapters 1 and 2 study statistical inference in a data structure that is common in the macroeconomic country-level studies (e.g., Acemoglu et al., 2008; Schularick and Taylor, 2012; Égert, 2016; Berka et al., 2018, analysed in Chapter 2), and applied experimental and behavioural literature (e.g., Gneezy et al., 2003; Reuben et al., 2017; Saccardo et al., 2018, mentioned in Chapter 3). The data sets in these studies are characterised by a relatively small number of cross-sectional units. The presence of heteroskedasticity and good leveraged data undermines conventional cluster-robust standard errors, leading to the over-rejection of the null hypothesis (Bramati and Croux, 2007). Chapter 3 performs an empirical exercise in an economic field – labour market segregation and working conditions – where it is problematic to claim for causality given self-selection into jobs (Petrongolo, 2004; Bertrand, 2011; Goldin, 2014; Bertrand, 2020; Morchio and Moser, 2021) and reverse causality (Borjas, 1980), making it difficult to disentangle the effect of segregation on employment contracts and wages. This thesis approaches these issues and proposes methods to address them.

In Chapter 1, I examined the statistical inference performed with conventional robust standard errors. In particular, I investigated the consequences of the simultaneous presence of small sample size, good leveraged data, and heteroskedastic disturbances on the validity of the statistical inference in linear panel data models. I formalised the panel versions of two jackknife-type estimators and proposed a new hybrid estimator. I derived their asymptotic distributions and analysed their finite sample properties in terms of proportional bias, rejection probability, root mean squared error, and adjusted power with Monte Carlo simulations. I found that test statistics obtained with conventional robust standard errors are over-sized, upward biased, and with low power under heteroskedasticity, with good leveraged data and in short panels. Under homoskedasticity and with good leveraged data, all estimators have good performances, suggesting that heteroskedasticity correction should be always used.

In Chapter 2, I developed diagnostic methods to detect three types of anomalous units (i.e., *good and bad leverage points*, and *vertical outliers*) in a panel data framework. I formalised statistical measures for quantifying the degree of leverage and outlyingness of units, and developed a method to visually detect the type of anomaly and its effect on other units. I first formalised the notion of the average individual leverage and average normalised residuals used for unit-wise leverage-residual plots. I proposed two diagnostic measures for the joint and conditional influence in a panel data setting. Then, I used network analysis tools to show the overall and bilateral influence exerted by pairs of units. I applied my method to a fictitious sample and to four country-level data sets used in published articles. I observed that bad and good leverage units have the largest joint and conditional influence and contribute to enhancing and masking the effects of even fairly influential units. Conversely, vertical outliers do not contribute in exerting large total influence.

Chapter 3, joint with Riccardo Leoncini and Mariele Macaluso, investigated the effect of gender sectoral segregation on employment contracts (part-time, permanent, remote work, number of weekly working hours) and hourly wages for both men and women. We first compared labour market outcomes of workers in female-dominated sectors with those in male-dominated sectors with similar observed socio-demographic and working characteristics using PSM. Then, we analysed the contribution of observable and unobservable factors in determining the gap in hourly wages using the Kitagawa-Blinder-Oaxaca decomposition and Mincerian wage regressions. We found that contractual features that are systematically chosen by a specific gender are more common in sectors dominated by that group for both genders. Workers employed in female-dominated sectors are on average paid less. The coefficient effect mainly explains the wage gap between men and women in both gender dominated sectors. Finally, we observed that women self-select into low-paid jobs, especially in female-dominated sectors.

This thesis is related to four strands of the literature. The first chapter created a link between the cross-sectional (Horn et al., 1975; MacKinnon and White, 1985; Davidson et al., 1993; Long and Ervin, 2000; Cribari-Neto, 2004; Cribari-Neto et al., 2007; Cribari-Neto and da Silva, 2011) and panel (Arellano, 1987) literature for Heteroskedasticity-Consistent (HC) estimators of the sampling variance. It supported the evidence that cluster-robust standard errors are downward biased in the presence of good leveraged data in the sample (Long and Ervin, 2000; Hayes and Cai, 2007; MacKinnon, 2013; Bramati and Croux, 2007; Verardi and Croux, 2009). In addition, this study – like Hinkley's (1977) – recommended the use of HC estimators because they are less sensitive to atypical cases, by construction. My contribution to the HC literature consisted in finding that conventional cluster-robust standard errors are always dominated by more conservative estimators of the variance, especially in short panels, and jackknife-type standard errors should be used.

Further, the second chapter of this thesis is related to the literature on diagnostic measures to detect atypical observations in the sample (Cook, 1979; Atkinson, 1985; Chatterjee and Hadi, 1988; Rousseeuw and Van Zomeren, 1990; Banerjee and Frees, 1997; Martín and Pardo, 2009; Martín, 2015; Pinho et al., 2015; Kim, 2017). More precisely I followed the school that recommends a local approach (Lawrance, 1995; Poon and Poon, 2001). This chapter extended Lawrance (1995) by developing diagnostic measures for the joint and conditional effects in panel data. The cross-sectional Cook's (1979) distance is an available statistical tool for the detection of unusual observations but fails to capture multiple anomalous cases because of the *masking effect*. By using a local approach, this chapter overcomes the issue faced with Cook-type measures.

In our third chapter, we referred to the matching literature to consistently estimate the average treatment effect in the absence of the treatment. We implemented matching estimators *à la* Abadie and Imbens (2006) that control for the propensity score eliminating the selection bias into jobs (Cameron and Trivedi, 2005, pp. 872-873).

Finally, our third study is located in a vast body of research on gender segregation at the workplace (Petrongolo, 2004; Mumford and Smith, 2008; Bertrand, 2011; Goldin, 2014; Bertrand, 2020; Morchio and Moser, 2021). While the gender discrimination literature has extensively documented that most of the pay gap comes from differences in human capital and estimated coefficients between men and women (Mumford and Smith, 2008), a part still remains unexplained and cannot be ascribed to only observed factors (Booth, 2009). We showed that women are negatively selected in terms of potential earnings such that differences in wage trajectories cannot be attributed to acquired skills and/or human capital only.

This thesis possesses three main novel elements. In Chapter 1, I proposed the use of three jackknife-type formulae for standard errors in place of conventional cluster-robust standard errors in a panel data setting. In Chapter 2, I exploited similarities between my data frame and the adjacency matrix of a weighted and directed graph. With this, I was able to mobilise network analysis tools to visually inspect the reciprocal influence exerted by units in the sample in a way that is more efficient than the traditional 2-way plots. Third, because the unexplained component after the KBO decomposition of wage differentials still remains unexplored in the gender economics literature, Chapter 3 delves into the unobservable factors using techniques from the economics of migration literature (Gould and Moav, 2016; Parey et al., 2017; Borjas et al., 2019).

Chapter 1

Robust Inference in Panel Data Models

Some Effects of Heteroskedasticity and Leveraged Data in Small Samples

1.1 Introduction

When the assumption of homoskedasticity is violated and the disturbances show non-constant variance (within i, over t, or both), least squares (LS) estimators are no longer efficient. Consequently, standard errors based on the incorrect assumption of homoskedastic disturbances lead to misleading statistical inferences. A common practice is to account for heteroskedasticity with robust standard errors when estimating the model. The Eicker (1967), Huber et al. (1967), and White's (1980) (henceforth, EHW) estimator has become the norm to account for any degree of heteroskedasticity in the cross-sectional environment. In the panel data structure, the Arellano's (1987) formula, based on EHW's estimator, has been widely used[1]. The presence of data points that exhibit extreme values in the covariates (referred as *good leverage points*) makes the EHW estimator systemati-

[1]We conducted a survey of articles published in the American Economic Review (AER) between 2011 and 2018 that make use of panel data estimation techniques. We observe that 206 out of 326 selected papers (65%) use linear least squares (LS) for panel data models – such as, Fixed-Effects (FE), First Differences (FD), Pooled Ordinary Least Squares (POLS), Linear Probability Model (LPM), Difference-in-Differences (DID), and Treatment Effects (TE) – as main (primary) estimation technique. Among this subgroup, 193 use cluster-robust standard errors, 15 use bootstrapped standard errors, and the remaining 36 use other types of standard errors. The strong preference for cluster-robust standard errors in our sample reflects the need to account for within-cluster correlation, heteroskedasticity and serial correlation whereas the choice of other types of standard errors handles other issues in the data structure. These figures are displayed in Figure C.1.

cally downward biased leading to liberal statistical inferences[2] (Long and Ervin, 2000; Godfrey, 2006; Hayes and Cai, 2007; MacKinnon, 2013; Şimşek and Orhan, 2016). In addition, the bias is severe when the cross-sectional sample size is sufficiently small[3] – i.e., the number of units is smaller than 250 (MacKinnon and White, 1985; Chesher and Jewitt, 1987; Silva, 2001; Verardi and Croux, 2009).

In some economic fields, researchers often face a data structure with rather small number of units, heteroskedasticity, and unusual data points. Such data structure is common in macroeconomic country-level analyses, and applied experimental and behavioural studies. Conventional cluster-robust standard errors are usually used to mainly control for within-group correlation (e.g., see replicated studies in Chapter 2) but this correction method is not appropriate because Arellano's (1987) standard errors are biased, having the same characteristics of the EHW estimator (Bramati and Croux, 2007; Verardi and Croux, 2009). While there is much discussion in the cross-sectional framework, little has been done for panel data[4]. In this chapter, we investigate the consequences of the simultaneous presence of small sample size, good leveraged data, and heteroskedastic disturbances on the validity of the statistical inference in linear panel data models.

We formalise panel versions of MacKinnon and White's (1985) and Davidson et al.'s (1993) estimators, and propose a new hybrid estimator, *PHC*6, that penalises only units with high leverage in the covariates. We derive the asymptotic distributions of a battery of estimators (i.e., Arellano's (1987), panel versions of MacKinnon and White's (1985) and Davidson et al.'s (1993), and PHC6 estimators), and analyse their finite sample properties with Monte Carlo (MC) simu-

[2] These anomalous cases do not affect the estimation of the least squares coefficients as shown in Figure C.6.

[3] This bias persists even in large samples.

[4] To the best of our knowledge, there are only two available studies for panel data. Kezdi (2003) compares the finite sample properties of a series of estimators of the variance-covariance matrix with an without serial correlation in the error term in large-N and small-T panels. Hansen (2007) derive the asymptotic properties of the conventional estimator of the variance-covariance matrix and studies its finite sample behaviour under heteroskedasticity in the cases where both N,T jointly go to infinity, and where either N or T goes to infinity holding the other dimension fixed. Extensions of a class of HC-based estimators to linear panel data mode ls has been conducted by Cattaneo et al. (2018) in high dimensional literature.

lations. Specifically, we compare the performances of these four types of estimators in terms of proportional bias, rejection probability (or empirical size), root mean squared error, and adjusted power. The analysis is conducted across different panel sample sizes and degrees of heteroskedasticity. Units are randomly contaminated with good leverage points. While we treat homoskedasticity as a special case, heteroskedasticity is assumed to be a core component of the correct regression specification. We focus on small sample sizes for a double reason. First, the cross-sectional HC literature has extensively discussed the finite sample bias of the EHW estimator and we suspect that the same bias also affects Arellano's (1987) estimator. Second, the nature of the research and/or data availability may force the investigator to deal with a reduced number of observations in the data set. Furthermore, we are interested in the additional issue posed by the presence of good leveraged data in this setting because they may be carried over the full history of a unit and, hence, contribute to exacerbate the effect on the estimates of the variance.

We find that under heteroskedasticity and with good leveraged data test statistics obtained with Arellano's (1987) standard errors are, as expected, over-sized, upward biased, and with low power, especially when the panel size is smaller than 2,500 observations. Test statistics calculated with PHC6 formula mimic the behaviour of those based on jackknife standard errors in terms of bias, empirical size and adjusted power test, converging to the same rates as the sample size increases. The panel version of MacKinnon and White's (1985) estimator shows similar patterns but with different magnitudes. Under homoskedasticity and with good leveraged data, all estimators have good performances in terms of proportional bias, rejection probabilities, and adjusted power, suggesting that the heteroskedasticity correction should be used. A similar result was found in MacKinnon and White (1985) and Long and Ervin (2000) for cross-sectional models who claimed that jackknife-type standard errors might enhance inference even with small degrees of heteroskedasticity.

Despite the remarkable methodological contribution in the cross-sectional HC literature, HC-type estimators – such as, HC2 by Horn et al. (1975), HC3 by MacKinnon and White (1985), HCjk by Davidson et al. (1993), HC4 by Cribari-Neto (2004), HC5 by Cribari-Neto et al. (2007), and HC4m by Cribari-Neto and da Silva (2011) – have not found much application in practice. However, this should not be a common practice because HC estimators alleviate the effect of leveraged data being less sensitive to anomalous cases, by construction (Hinkley, 1977). This study contributes to the HC literature by creating a link between cross-sectional and panel HC estimators of the sampling variance. We provide the formulae and derive the distribution of a selected group of variance-covariance estimators to panel data. We document the downward bias of conventional robust standard errors under certain circumstances and provide alternative solutions to obtain more reliable statistical inferences. This study provides simulation evidence that these estimators outperform the conventional cluster-robust standard errors under specific circumstances and should be used in linear panel data models.

The rest of the chapter is structured as follows. Section 1.2 introduces the static linear panel data model and its assumptions. Sections 1.3 and 1.4 discuss the asymptotic properties of the estimators. In Section 1.5, we extend a selected group of HC estimators to panel data and propose a new estimator. Then, Section 1.7 shows the MC simulation design and discusses the simulation results. In Section 1.8, we examine the performances of the four estimators in terms of their proportional bias, empirical size, adjusted power, and mean squared errors. Section 3.5 concludes.

1.2 Model and Assumptions

Consider the static linear panel regression model with one-way error component

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \alpha_i + u_{it}, \;\; i \in \mathscr{I} = \{1, \dots, N\} \text{ and } t \in \mathscr{T} = \{1, \dots, T\} \tag{1.1}$$

where y_{it} is the response variable for the cross-sectional unit i at time period t; $\mathbf{x}_{it}$ is a $k \times 1$ vector of time-varying inputs, $\boldsymbol{\beta}$ is a $k \times 1$ vector of parameters of interest; α_i is the individual-specific unobserved heterogeneity (or *fixed effects*); and u_{it} is a stochastic error component.

Stacking observations for t, model (1.1) at the level of the observation becomes

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\alpha}_i + \mathbf{u}_i, \text{ for all } i = 1, \dots, N, \tag{1.2}$$

where $\mathbf{y}_i$ is $T \times 1$ vector of outcomes; $\mathbf{X}_i$ is a $T \times k$ matrix of time-varying regressors; $\boldsymbol{\alpha}_i = \alpha_i \boldsymbol{\iota}$ is a $T \times 1$ vector of individual fixed effects, and $\boldsymbol{\iota}$ is a vector of ones of order T; and $\mathbf{u}_i$ is a $T \times 1$ vector of one-way error component. The fixed effects $\boldsymbol{\alpha}_i$ in Equation (2.1) are removed to consistently estimating the parameter of interest $\boldsymbol{\beta}$ by applying an appropriate transformation of the original data, i.e., the *time-demeaning* or *first-differencing* procedure, because it might be the case that $\mathbb{E}(\boldsymbol{\alpha}_i|\mathbf{X}_i) = h(\mathbf{X}_i)$. For the rest of the discussion, we focus on the first approach when applied to Equation (2.1). The *time-demeaning* data transformation delivers a consistent estimator of $\boldsymbol{\beta}$ even when the regressor is correlated with the unobserved heterogeneity α_i, but is less efficient than the First-Difference (FD) transformation with errors that are not identically distributed.

The estimating equation becomes

$$\widetilde{\mathbf{y}}_i = \widetilde{\mathbf{X}}_i\boldsymbol{\beta} + \widetilde{\mathbf{u}}_i, \text{ for all } i = 1, \dots, N, \tag{1.3}$$

where $\widetilde{\mathbf{y}}_i = (\mathbf{I}_T - T^{-1}\boldsymbol{\iota\iota}')\mathbf{y}_i$ is $T \times 1$; $\widetilde{\mathbf{X}}_i = (\mathbf{I}_T - T^{-1}\boldsymbol{\iota\iota}')\mathbf{X}_i$ is $T \times k$; and $\widetilde{\mathbf{u}}_i = (\mathbf{I}_T - T^{-1}\boldsymbol{\iota\iota}')\mathbf{u}_i$ is $T \times 1$. Note that $(\mathbf{I}_T - T^{-1}\boldsymbol{\iota\iota}')\boldsymbol{\alpha}_i = \mathbf{0}$ as $T^{-1}\boldsymbol{\iota\iota}'\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_i$. The within-group estimator is the Pooled OLS estimator of Equation (1.3).

The model assumptions are as follows

ASM.1 (*data-generating process*):

i (*independent variables*): $\{\mathbf{X}_i\}$ is an independent and identically distributed (*iid*) sequence of random variables, for all $i = 1, \ldots, N$;

ii (*disturbances*): $\{\mathbf{u}_i\}$ is an independent but not identically distributed (*inid*) sequence of random error terms, for all $i = 1, \ldots, N$.

ASM.2 (*on the relation of* $\widetilde{\mathbf{X}}_i$ *and* $\widetilde{\mathbf{u}}_i$):

i (*strong exogeneity*): $\mathbb{E}\left(\widetilde{\mathbf{u}}_i|\widetilde{\mathbf{X}}_i\right) = 0$, for all $i = 1, \ldots, N$;

ii (*heteroskedasticity*): $\overline{\boldsymbol{\Sigma}}_N = N^{-1}\sum_{i=1}^{N}\boldsymbol{\Sigma}_i \to \boldsymbol{\Sigma}$, where the matrix of the heteroskedastic disturbances $\boldsymbol{\Sigma}_i = \mathbb{E}\left(\widetilde{\mathbf{u}}_i\widetilde{\mathbf{u}}_i'|\widetilde{\mathbf{X}}_i\right) = \text{diag}\left\{\sigma_{it}^2\right\}$ is symmetric of dimension T, finite, positive definite, and diagonal.

The above model assumptions have the following implications. ASM.1.i guarantees that the sequence of random variables $\{\mathbf{X}_i'\mathbf{X}_i\}$ is *iid* [PROP 3.3 in White (1984, p.30)]. ASM.1.ii imposes cross-sectional independence and, together with ASM.1.i, implies that $\{\mathbf{X}_i'\mathbf{u}_i\}$ is an *inid* sequence of random vectors [PROP 3.10 in White (1984, p.34)]. Assumption ASM.1 and its implications remain unaltered after any data transformation.

The strict exogeneity assumption ASM.2.i rules out feedback effects and implies contemporaneous exogeneity, i.e., $\mathbb{E}\left(\widetilde{u}_{it}|\widetilde{\mathbf{X}}_i\right) = \mathbb{E}\left(\widetilde{u}_{it}|\widetilde{\mathbf{x}}_{it}\right) = 0$, and is a crucial assumption to prove consistency of the *within-group* estimator. The projection analog of ASM.2.i is the strong exogeneity condition, i.e, $\mathbb{E}\left(\widetilde{\mathbf{x}}_{is}\widetilde{u}_{it}\right) = 0 \Leftrightarrow \mathbb{E}\left(\widetilde{u}_{it}|\widetilde{\mathbf{X}}_i\right)$=0, for all $s \in \mathscr{T}$ and $s \neq t$. Because the exogeneity of the non-demeaned variables might not be strong enough to guarantee that exogeneity is preserved *after* the transformation[5], i.e., $\mathbb{E}\left(\mathbf{X}_i'\mathbf{u}_i\right)$=0 $\not\Rightarrow \mathbb{E}\left(\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{u}}_i\right)$=0 (Cameron and Trivedi, 2005, p.707). Assumption ASM.2.ii allows the conditional error variance to vary across observations and time periods, and imposes serial uncorrelation over time dimension, $\mathbb{E}\left(\widetilde{u}_{it}\widetilde{u}_{is}|\widetilde{\mathbf{X}}_i\right) = 0$ with $(t,s) \in \mathscr{T}$ and $t \neq s$.

The assumptions for the existence and optimality properties of the estimator of the true population parameter $\boldsymbol{\beta}$ are

[5] This occurs because the regressor is correlated with $T^{-1}\boldsymbol{\iota}\boldsymbol{\iota}'\widetilde{\mathbf{u}}_i$ since it includes the whole history.

ASM.3 (*rank condition*): $\mathbf{S}_N \equiv N^{-1}\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{X}}_i$ is a finite symmetric matrix with full column rank k.

ASM.4 (*moment conditions*):

i $\mathbb{E}\left\|\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{X}}_i\right\| < \infty$ for $\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{X}}_i \in \mathbb{R}^{k\times k}$;

ii $\sup_i \mathbb{E}\left\|\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{u}}_i\right\|^{2+\delta} < \infty$ for some $\delta > 0$, $\forall i$ and $\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{u}}_i \in \mathbb{R}^k$,

where $\|\cdot\|$ denotes the Euclidean norm.

ASM.5 (*average variance-covariance matrix convergence*):

$\overline{\mathbf{V}}_N = N^{-1}\sum_{i=1}^{N}\mathbf{V}_i \to \mathbf{V}$, where $\mathbf{V}_i = \mathbb{E}\left(\widetilde{\mathbf{X}}_i'\Sigma_i\widetilde{\mathbf{X}}_i\right)$ and $\mathbf{V}$ is a finite positive definite $k\times k$ matrix.

The full column rank condition in ASM.3 implies non-singularity of the matrix $\mathbf{S}_N$ and, hence, no perfect multicollinearity that guarantees the invertibility of the matrix. The limiting matrix $\mathbf{S}_{XX} \equiv \mathbb{E}\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)$ possesses the properties of $\mathbf{S}_N$ by the *Weak Law of Large Numbers* (*WLLN*) [THM 6.6]. Another implication of ASM.3 is that the matrix of regressors $\widetilde{\mathbf{X}}_i$ is full column rank. Assumption ASM.4 defines the finiteness and boundedness of moments in terms of the Euclidean norm. Assumption ASM.5 ensures that the average variance-covariance matrix converges to a finite quantity, satisfying one of the conditions of the *Multivariate Central Limit Theorem* (*MCLT*) for *inid* processes [THM 6.16 in Hansen (2019, p.189)].

No restrictions are placed on influential points – i.e., high leverage points and outliers[6] – possibly allowing for their presence. We consider a framework where the panel is small, that is, the time period length is smaller than the number of units N such that $T \ll N$. Under this notation T is the full set of time information, and the total number of observations in the sample is given by $n = N\cdot T$ with balanced data sets.

[6]Anomalous observations that have extreme values in the covariate are high leverage points whereas outliers are points with large residuals (Rousseeuw and Van Zomeren, 1990; Chatterjee and Hadi, 1986). See Chapter 2 for further discussion.

This set of assumptions and their implications remain valid under any monotonic data transformation due to the *Continuous Mapping Theorem* (*CMT*) [THM 6.19 in Hansen (2019, p.192)]. Later in this work, we consider the *within-group* transformation of the data.

1.3 Asymptotic Properties of β

Under ASM.1–ASM.4.i, the *within-group* estimator of the true population parameter $\boldsymbol{\beta}$ exists with form $\widehat{\boldsymbol{\beta}}_N = \left(N^{-1}\sum_{i=1}^N \widetilde{\mathbf{X}}_i'\widetilde{\mathbf{X}}_i\right)^{-1} N^{-1}\sum_{i=1}^N \widetilde{\mathbf{X}}_i'\widetilde{\mathbf{y}}_i$, and is consistent, i.e.,

$$\widehat{\boldsymbol{\beta}}_N - \boldsymbol{\beta} = \left(\frac{1}{N}\sum_{i=1}^N \widetilde{\mathbf{X}}_i'\widetilde{\mathbf{X}}_i\right)^{-1} \frac{1}{N}\sum_{i=1}^N \widetilde{\mathbf{X}}_i'\widetilde{\mathbf{u}}_i \xrightarrow{p} \mathbf{0} \text{ as } N \to \infty. \tag{1.4}$$

The consistency of the *within-group* estimator under the aforementioned assumptions is a known result (as reference, see Hansen, 2019, pp. 612–613). By the previously discussed implication of ASM.1.i and PROP 3.3 in White (1984, p.30), $\left\{\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{X}}_i\right\}$ is an *iid* sequence of random variables with finite moments given ASM.4. The elements of the sequence satisfy the *Weak Law of Large Numbers* (*WLLN*) [THM 6.6 in Hansen (2019, p.182)] such that $\mathbf{S}_N \xrightarrow{p} \mathbf{S}_{XX} < \infty$. Because both matrices are invertible by ASM.3, then THM 6.19 [*Continuous Mapping Theorem (CMT)* in Hansen (2019, p.192)] yields the result $\mathbf{S}_N^{-1} \xrightarrow{p} \mathbf{S}_{XX}^{-1}$.

Now, we show that the second component in (1.4) converges in probability to zero. We know that the sequence $\left\{\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{u}}_i\right\}$ is *inid* as an implication of ASM.1 [PROP 3.10 in White (1984, p.33)]. Then, the Chebyshev inequality is

$$\Pr\left(\left\|\frac{1}{N}\sum_{i=1}^N \widetilde{\mathbf{X}}_i'\widetilde{\mathbf{u}}_i\right\| \geq \varepsilon\right) \leq \frac{\mathbb{E}\left\|\frac{1}{N}\sum_{i=1}^N \widetilde{\mathbf{X}}_i'\widetilde{\mathbf{u}}_i\right\|^2}{\varepsilon^2}, \tag{1.5}$$

where the numerator in (1.5) can be expanded as follows

$$\begin{aligned}\left\|\frac{1}{N}\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{u}}_i\right\|^2 &= \operatorname{tr}\left\{\left(\frac{1}{N}\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{u}}_i\right)\left(\frac{1}{N}\sum_{j=1}^{N}\widetilde{\mathbf{u}}_j'\widetilde{\mathbf{X}}_j\right)\right\} \\ &= \frac{1}{N^2}\operatorname{tr}\left\{\sum_i\sum_j\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{u}}_i\widetilde{\mathbf{u}}_j'\widetilde{\mathbf{X}}_j\right\}.\end{aligned} \tag{1.6}$$

By the aforementioned implication of ASM.1.ii and under ASM.2.ii the conditional error variance is

$$\mathbb{E}\left(\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{u}}_i\widetilde{\mathbf{u}}_j'\widetilde{\mathbf{X}}_j\middle|\widetilde{\mathbf{X}}_i\widetilde{\mathbf{X}}_j\right) = \begin{cases}\mathbf{0} & \forall i \neq j \\ \widetilde{\mathbf{X}}_i'\Sigma_i\widetilde{\mathbf{X}}_i & \forall i = j\end{cases} \tag{1.7}$$

Applying the expected value operator to (1.6), and using result (1.7) jointly with the *Law of Iterated Expectations* (LIE), the above equality becomes as follows

$$\begin{aligned}\mathbb{E}\left\|\frac{1}{N}\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{u}}_i\right\|^2 &= \frac{1}{N^2}\operatorname{tr}\left\{\sum_i\sum_j\mathbb{E}\left(\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{u}}_i\widetilde{\mathbf{u}}_j'\widetilde{\mathbf{X}}_j\right)\right\} \\ &= \frac{1}{N^2}\operatorname{tr}\left\{\sum_i\sum_j\mathbb{E}\left[\mathbb{E}\left(\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{u}}_i\widetilde{\mathbf{u}}_j'\widetilde{\mathbf{X}}_j\middle|\widetilde{\mathbf{X}}_i,\widetilde{\mathbf{X}}_j\right)\right]\right\} \\ &= \frac{1}{N^2}\operatorname{tr}\left\{\sum_i\mathbb{E}\left[\widetilde{\mathbf{X}}_i'\Sigma_i\widetilde{\mathbf{X}}_i\right]\right\} \\ &= \frac{1}{N}\operatorname{tr}\left\{\frac{1}{N}\sum_i\mathbb{E}\left[\widetilde{\mathbf{X}}_i'\Sigma_i\widetilde{\mathbf{X}}_i\right]\right\} \\ &= \frac{1}{N}\operatorname{tr}\left\{\overline{\mathbf{V}}_N\right\} \to 0, \text{ as } N \to \infty\end{aligned} \tag{1.8}$$

since assumption (ASM.5) implies that $\operatorname{tr}\left\{\overline{\mathbf{V}}_N\right\} \to \operatorname{tr}\left\{\mathbf{V}\right\}$, which is finite.

As a result, the right-hand side of Equality (1.8) converges in probability to zero. So does the left-hand side. Inequality (1.5) becomes

$$\Pr\left(\left\|\frac{1}{N}\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{u}}_i\right\| \geq \varepsilon\right) \to 0 \text{ as } N \to \infty,$$

and, hence, $N^{-1}\sum_i \widetilde{\mathbf{X}}_i'\widetilde{\mathbf{u}}_i \xrightarrow{p} \mathbf{0}$. By THM 6.19 [CMT in Hansen (2019, p.192)] , the result follows $\widehat{\boldsymbol{\beta}}_N - \boldsymbol{\beta} \xrightarrow{p} \mathbf{S}_{XX}^{-1}\cdot\mathbf{0} = \mathbf{0}$, or alternatively $\widehat{\boldsymbol{\beta}}_N \xrightarrow{p} \boldsymbol{\beta}$. This result holds for any monotonic transformation of the data.

Under ASM.1–ASM.5, the estimator has the known asymptotic distribution below

$$\sqrt{N}\big(\widehat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}\big) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{S}_{XX}^{-1}\mathbf{V}\mathbf{S}_{XX}^{-1}\right) \quad \text{as } N \to \infty \text{ and } T \text{ fixed.} \tag{1.9}$$

A reference for this result is Hansen (2019, pp. 624–625). The left-hand-side of Equation (1.9) can be re-written as follows

$$\sqrt{N}\big(\widehat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}\big) = \left(\frac{1}{N}\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{X}}_i\right)^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{u}}_i.$$

The sequence of random variables $\left\{\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{X}}_i\right\}$ is *iid* as implication of ASM.1.i and by PROP 3.3 in White (1984, p.30). With analogous arguments as those used above to prove consistency, $\mathbf{S}_N^{-1} \xrightarrow{p} \mathbf{S}_{XX}^{-1}$. Under assumptions ASM.1 and ASM.2.i and by PROP 3.10 in White (1984, p.33), the sequence $\left\{\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{u}}_i\right\} \in \mathbb{R}^k$ is *inid* with means $\mathbb{E}\big(\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{u}}_i\big) = 0$ and variance matrices $\mathbf{V}_i = \mathbb{E}\big(\widetilde{\mathbf{X}}_i'\boldsymbol{\Sigma}_i\widetilde{\mathbf{X}}_i\big)$, by LIE and ASM.2.ii. The limit in probability ASM.5 and assumption ASM.4.ii are the two conditions that satisfy the *Multivariate Central Limit Theorem* (*MCLT*) for *inid* processes [THM 6.16 in Hansen (2019, p.189)]. Therefore, $N^{-1/2}\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{u}}_i \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{V}\right)$ as $N \to \infty$. *Slutsky's Theorem* [THM 6.22.2 in Hansen (2019, p.193)] yields the result $\sqrt{N}\big(\widehat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}\big) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{S}_{XX}^{-1}\mathbf{V}\mathbf{S}_{XX}^{-1}\right)$, where $\mathbf{S}_{XX} \equiv \mathbb{E}\big(\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{X}}_i\big)$. THM 6.19 [Hansen (2019, p.192)] ensures that the above limits hold for any monotonic transformation of the data, e.g., the within-group transformation.

1.4 Estimating the Asymptotic Variance

Given the above results under the model assumptions we made, the approximate distribution of the estimator of $\boldsymbol{\beta}$ for large but finite samples is

$$\widehat{\boldsymbol{\beta}}_N \overset{a}{\sim} \mathcal{N}(\boldsymbol{\beta}, N^{-1}\mathbf{S}_{XX}^{-1}\mathbf{V}\mathbf{S}_{XX}^{-1}), \tag{1.10}$$

where the limiting matrices $\mathbf{S}_{XX}$ and $\mathbf{V}$ need to be estimated, and so does the average variance-covariance matrix $\overline{\mathbf{V}}_N$. While $\mathbf{S}_{XX}$ is estimated by $\mathbf{S}_N = N^{-1}\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{X}}_i$, the estimation of the average variance-covariance matrix needs further discussion. According to White (1980), a computationally feasible practice consists in estimating each expectation, $\mathbf{V}_i = \mathbb{E}\big(\widetilde{\mathbf{X}}_i'\Sigma_i\widetilde{\mathbf{X}}_i\big)$, individually, and a plausible estimator of $\overline{\mathbf{V}}_N$ would be $N^{-1}\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{u}}_i\widetilde{\mathbf{u}}_i'\widetilde{\mathbf{X}}_i$ if the error term were known. Because it is unobserved, a consistent estimator of the variance-covariance matrix is in practice $N^{-1}\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'\widehat{\widetilde{\mathbf{u}}}_i\widehat{\widetilde{\mathbf{u}}}_i'\widetilde{\mathbf{X}}_i$, where $\widehat{\widetilde{\mathbf{u}}}_i = \widetilde{\mathbf{y}}_i - \widetilde{\mathbf{X}}_i\widehat{\boldsymbol{\beta}}$. Define $\widehat{\widetilde{\mathbf{u}}}_i = \widehat{\mathbf{u}}_i$ to simplify the notation.

Using a generalised expression for regression residuals, the variance-covariance matrix can be re-written as follows: $\widehat{\mathbf{V}}_N = N^{-1}\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'\widehat{\mathbf{v}}_i\widehat{\mathbf{v}}_i'\widetilde{\mathbf{X}}_i$, where $\widehat{\mathbf{v}}_i = \mathbf{M}_i^{-1}\widehat{\widetilde{\mathbf{u}}}_i$ are the transformed regression residuals with $\mathbf{M}_i$ being the transformation matrix that differs across estimators of the variance-covariance. When the transformed residuals equalise the residuals from the regression, $\widehat{\mathbf{v}}_i = \mathbf{I}_T\widehat{\mathbf{u}}_i$, the variance-covariance matrix takes the familiar *"sandwich-like"* formula of Arellano's (1987) estimator. Other functional forms of the transformed residuals are presented and discussed in Section 1.5. In addition, the consistency of transformed residuals is proved under fixed T in Appendix A.3.

The variance-covariance matrix $\widehat{\mathbf{V}}_N$ with transformed residuals is still a consistent estimator of the true variance. Let $\widehat{\boldsymbol{\Sigma}}_i = \widehat{\mathbf{v}}_i\widehat{\mathbf{v}}_i'$, from White's (1980) general result and under the above model assumptions and THM 7.7 in Hansen (2019, p.232), it follows that $\big\|\widehat{\mathbf{V}}_N - \overline{\mathbf{V}}_N\big\| \xrightarrow{p} \mathbf{0}$ and, hence, $\left\|N^{-1}\sum_i \widehat{\Sigma}_i - \overline{\Sigma}_N\right\| \xrightarrow{p} \mathbf{0}$, for all $i = 1,\ldots,N$, as $N \to \infty$ and keeping T fixed.

1.5 Estimators of the Variance for Panel Data

The EHW estimator and the other HC estimators of the variance were originally formalised for cross-sectional models. In this section, we present the formulae of a battery of HC estimators for panel data, alongside with Arellano's (1987) well-known formula. We formalise MacKinnon and White's (1985) jackknife-type estimator for panel data, provide a panel version of Davidson et al.'s (1993) estimator, and propose a new hybrid estimator, *PHC*6.

1.6 HC*k*-type Estimators

The well-known formula of Arellano's (1987) estimator (henceforth, PHC0) is

$$\widehat{\text{AVar}(\widehat{\beta})}_0 = c_0\,\mathbf{S}_N^{-1}\widehat{\mathbf{V}}_N^0\mathbf{S}_N^{-1}\,, \tag{1.11}$$

where $c_0 = \frac{n-1}{n-k}\cdot\frac{N}{N-1}$, and $\widehat{\mathbf{V}}_N^0 = N^{-1}\sum_{i=1}^N \widetilde{\mathbf{X}}_i'\widehat{\mathbf{v}}_i\widehat{\mathbf{v}}_i'\widetilde{\mathbf{X}}_i$ with $\mathbf{M}_i = \mathbf{I}_T$. The finite-sample correction factor[7], c_0, ensures that $\widehat{\mathbf{V}}_N^0$ is consistent under ASM.2.ii with fixed T; the ratio $N/(N-1)$ is a computational necessary degree-of-freedom correction to control for individual correlation (Stock and Watson, 2008; Cameron et al., 2011).

The estimator that resembles Davidson et al.'s (1993) HC3 in the panel data framework (PHC3) is as follows

$$\widehat{\text{AVar}(\widehat{\beta})}_3 = c_3\,\mathbf{S}_N^{-1}\widehat{\mathbf{V}}_N^3\mathbf{S}_N^{-1}\,, \tag{1.12}$$

[7]Computationally, statistical software, like STATA, use a finite-sample modification of the conventional (i.e., Arellano's (1987)) variance-covariance matrix multiplying $N^{-1}\sum_{i=1}^N \widetilde{\mathbf{X}}_i'\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\widetilde{\mathbf{X}}_i$ by the correction factor $c = \frac{n-1}{n-k}\cdot\frac{N}{N-1}$, where $n = N\cdot T$ for one-way clustering in panel data, otherwise cluster-robust standard error turn out to be downward biased (Arellano, 1987; Bertrand et al., 2004; Cameron et al., 2011).

where $c_3 = (N-1)N^{-1}$, $\widehat{\mathbf{V}}_N^3 = \frac{1}{N}\sum_{i=1}^N \widetilde{\mathbf{X}}_i'\widehat{\mathbf{v}}_i\widehat{\mathbf{v}}_i'\widetilde{\mathbf{X}}_i$ with $\mathbf{M}_i = (\mathbf{I}_T - \mathbf{H}_i)$ and the individual leverage matrix[8], $\mathbf{H}_i = \widetilde{\mathbf{X}}_i\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}_i'$, whose diagonal elements $h_{itt} = \widetilde{\mathbf{x}}_{it}'(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{x}}_{it}$ lie in the $(0,1)$ interval but the off-diagonal elements may be negative. Predicted residuals, $\widehat{\mathbf{v}}_i$, assign a penalty to LS residuals based on the degree of leverage making the estimates of the variance less sensitive to leverage points. This type of standard errors tend to be asymptotically conservative as the number of covariates is allowed to grow as fast as the sample size, despite being asymptotically valid (Cattaneo et al., 2018).

The estimator of the jackknife asymptotic variance for panel data models (PHCjk) adapts MacKinnon and White's (1985) HCjk estimator and has form

$$\begin{aligned}\widehat{\text{AVar}(\widehat{\boldsymbol{\beta}})}_{jk} &= \left(\frac{N-1}{N}\right)\sum_{i=1}^{N}\left(\widehat{\boldsymbol{\beta}}_{(i)} - \bar{\boldsymbol{\beta}}\right)\left(\widehat{\boldsymbol{\beta}}_{(i)} - \bar{\boldsymbol{\beta}}\right)' \\ &= \left(\frac{N-1}{N^2}\right)\mathbf{S}_N^{-1}\left\{\widehat{\mathbf{V}}_N^3 - \boldsymbol{\mu}^*\boldsymbol{\mu}^{*\prime}\right\}\mathbf{S}_N^{-1},\end{aligned} \tag{1.14}$$

where the Leave-One-Out estimator is $\widehat{\boldsymbol{\beta}}_{(i)} = \widehat{\boldsymbol{\beta}} - \left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}_i'\widehat{\mathbf{v}}_i$ with $\mathbf{M}_i = (\mathbf{I}_T - \mathbf{H}_i)$, $\bar{\boldsymbol{\beta}} = \frac{1}{N}\sum_{i=1}^N \widehat{\boldsymbol{\beta}}_{(i)}$, and $\boldsymbol{\mu}^* = \frac{1}{N}\sum_{i=1}^N \widetilde{\mathbf{X}}_i'\widehat{\mathbf{v}}_i$. The jackknifed variance-covariance estimator with *fixed effects* can be found in Belotti and Peracchi (2020).

In practice, the jackknife procedure consists in deleting the entire history of each unit one at a time without replacement. Because the jackknife resamples in such a way to construct "pseudo-data" on which the estimator of interest is tested, this technique – as well as the bootstrap – is suitable for the assessment of the variability of an estimate, e.g., the estimation of standard errors (Efron, 1982; Freedman and Peters, 1984, Chapter 6). The advantages of the jackknife

[8]The individual leverage matrix is a $T \times T$ matrix defined as follows

$$\mathbf{H}_i = \begin{pmatrix} h_{i11} & h_{i12} & \dots & h_{i1T} \\ h_{i21} & h_{i22} & \dots & h_{i21T} \\ \vdots & \vdots & \vdots & \vdots \\ h_{iT1} & h_{iT2} & \dots & h_{iTT} \end{pmatrix} \quad \text{for all } i = 1,\dots,N \tag{1.13}$$

with elements $h_{its} = \widetilde{\mathbf{x}}_{it}'(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{x}}_{is}$ with $t,s = 1,\dots,T$.

procedure are double: it is an entirely data-driven approach, and it is able to alleviate the impact of influential units on inference (Cattaneo et al., 2019). The main drawback is that the jackknife estimator becomes computationally infeasible for sufficiently large number of groups.

PHC3 is a special case of Equation (1.14) when the contribution of the second block is null as $N \to \infty$ and fixing T. The two estimators are asymptotically equivalent and coincide in sufficiently large samples. The derivation of (1.14) involves considerable algebraic manipulations (see Appendix A.1).

1.6.1 A Hybrid Estimator: PHC6

We propose a hybrid estimator of the variance, PHC6, that nests PHC0 and PHC3 estimators using a threshold criterion from the decision rule of the penalty factor in Cribari-Neto (2004). PHC3 is chosen because Monte Carlo simulations showed that Davidson et al.'s (1993) HC3 possess the best final sample properties in terms of lower bias, with rejection rates closer to the nominal one (Long and Ervin, 2000). The threshold criterion is designed to account for the time period in which each unit has exerted the maximal leverage with respect to the average leverage in the same period. PHC6 is designed to deliver standard errors that are higher in magnitude than PHC0 with contaminated observations but the same as PHC0 standard errors with no extreme observations in the sample.

Before presenting the proposed estimator, we clarify beforehand the notation we will be using. Let the $T \times 1$ vector

$$\mathbf{h}_i = diag(\mathbf{H}_i) = \begin{pmatrix} h_{i11} \\ h_{i22} \\ \vdots \\ h_{iTT} \end{pmatrix} \text{ for all } i = 1, \dots, N$$

be the individual leverage vector constructed from the diagonal elements of the individual leverage matrix $\mathbf{H}_i$ defined in (1.13), and let the $T \times 1$ vector

$$\overline{\mathbf{h}}_{tt} = \begin{pmatrix} \overline{h}_{11} \\ \overline{h}_{22} \\ \vdots \\ \overline{h}_{TT} \end{pmatrix} = \begin{pmatrix} N^{-1}\sum_{i=i}^{N} h_{i11} \\ N^{-1}\sum_{i=i}^{N} h_{i22} \\ \vdots \\ N^{-1}\sum_{i=i}^{N} h_{iTT} \end{pmatrix}$$

be constructed from the average leverage at time t across units. Then, let $\overline{\mathbf{h}}$ be a $T \times 1$ vector with elements $\left(\overline{\mathbf{h}}_{tt} \exp \circ \mathbf{j}\right)$, where the expression $\exp \circ \mathbf{j}$ indicates the element-wise power of $\mathbf{j}$ which is a $T \times 1$ vector of negative ones. The Hadamard (element-wise) product, $\mathbf{h}_i \odot \overline{\mathbf{h}}$, is a $T \times 1$ vector whose elements, $h_{itt}\overline{h}_{tt}^{-1}$, inform on the relative leverage of unit i at time t with respect to the average leverage at time t. Specifically, values of $h_{itt}\overline{h}_{tt}^{-1}$ above one signal that the relative leverage of unit i at time t exceeds the average influence at time t. Units with values slightly grater than one cannot automatically be flagged as highly influential because in the absence of influential units at time t, the denominator may be very close to the numerator, by construction and, hence, one cannot be chosen as cut-off value. Conversely, high values of $h_{itt}\overline{h}_{tt}^{-1}$ indicate that unit i is exerting high leverage at time t with respect to the mean influence at time t.

The PHC6 estimator of the variance is defined as follows

$$\widehat{\text{AVar}(\widehat{\boldsymbol{\beta}})}_6 = c_6 \mathbf{S}_N^{-1} \widehat{\mathbf{V}}_N^6 \mathbf{S}_N^{-1}, \tag{1.15}$$

where the variance-covariance matrix is $\widehat{\mathbf{V}}_N^6 = \frac{1}{N}\sum_{i=1}^{N} \widetilde{\mathbf{X}}_i' \widehat{\mathbf{v}}_i \widehat{\mathbf{v}}_i' \widetilde{\mathbf{X}}_i$, and the matrix $\mathbf{M}_i$ has functional form

$$\mathbf{M}_i = \begin{cases} \mathbf{I}_T & \text{if } h_i^* < 2 \\ \mathbf{I}_T - \mathbf{H}_i & \text{otherwise} \end{cases} \tag{1.16}$$

where $h_i^* = max\{h_{i11}/\overline{h}_{11}, \ldots, h_{iTT}/\overline{h}_{TT}\}$ is the maximal individual leverage of unit i; and $\overline{h}_{tt} = N^{-1}\sum_{i=i}^{N} h_{itt}$ is the average leverage at time t, with h_{itt} being the individual leverage of unit i at time t. The finite sample correction of PHC6 is

$$c_6 = \begin{cases} \frac{(NT-1)N}{(NT-k)(N-1)} & \text{if } h_i^* < 2 \\ \frac{N-1}{N} & \text{otherwise} \end{cases}.$$

According to the cut-off rule, residuals of units with maximal individual relative leverage, $h_i^* = h_{itt}\overline{h}_{tt}^{-1}$, are discounted by the penalty matrix $\mathbf{M}_i$. Unlike PHC3 that penalises both low and high leverage points at the same rate, PHC6 discounts at the same discounting rate as PHC3 only if the unit exerts high leverage. When the individual relative leverage does not exceed the cutoff, no penalty is applied and PHC6 coincides with Arellano's (1987) estimator. Conversely, when the average level of leverage exceeds the cut-off value, PHC6 residuals are penalised as in PHC3. In addition, PHC6 always weights for a final sample correction.

The cut-off is set to be equal to 2 such that no penalty is assigned to fairly influential units at time t. One is not chosen as a cutoff value because in the absence of anomalous cases, the denominator $\overline{h}_{tt}$ would be very close to the numerator h_{itt} for some units with meaningless individual leverage but above the mean average. This would drive the ratio to exceed one.

1.7 Monte Carlo Simulation

In this section, we present the MC simulation design which illustrates the behaviour of the four types of estimators of the variance in finite samples[9], when variables are contaminated with anomalous data points. For simplicity, the simulation set up uses synthetic balanced data set and does not allow for any correlation between the individual-specific fixed effects and the regressor[10]. The data generating process for the Monte Carlo simulation is designed to be closely related to: (i) Godfrey (2006), Stock and Watson (2008), and MacKinnon (2013) in terms of the form of heteroskedasticity, number of regressors and the calibrated parameters; and (ii) Bramati and Croux (2007) for the contamination with cell-isolated good leverage points. However, we depart from these settings by making some modifications to the simulation designs.

The data generating process (DGP) of Monte Carlo simulations is as follows

$$y_{it} = \beta_0 + \sum_{k=1}^{K} \beta_j x_{it,k} + \alpha_i + u_{it}, \text{ for all } i = 1,\dots,N \text{ and } t = 1,\dots,T_i \tag{1.17}$$

$$x_k \sim \mathcal{N}(0,1) \text{ for } k = \{1,2\} \text{ except contaminated cases} \tag{1.18}$$

$$x_k = f(x_1,x_2) \text{ for } k = \{3,4,5\} \tag{1.19}$$

$$\alpha_i \sim U(0,1) \tag{1.20}$$

$$u_{it} = \sigma_{it}\varepsilon_{it} + \theta\varepsilon_{it-1},\ \varepsilon_{it} \sim \mathcal{N}(0,1),\ u_{it} \sim \mathcal{N}(0,\sigma_{it}^2) \tag{1.21}$$

$$\sigma_{it}^2 = z(\gamma)\left(\beta_0 + \sum_{j=1}^{J} \beta_j x_{it,j}\right)^{\gamma}, \text{ with } z(\gamma) = \left[\mathbb{E}\left(\beta_0 + \sum_{j=1}^{J} \beta_j x_{it,j}\right)^{\gamma}\right]^{-1} \tag{1.22}$$

where the number of regressors in the model is $K = 5$ and $K = J$; model parameters are calibrated to

[9]Monte Carlo simulations provide computational evidence of finite sample properties of an estimator or a test when applied to fictitious data (Hendry, 1984; Kiviet et al., 2012).

[10]This design leaves open the possibility to estimate the regression equation consistently and efficiently using the *random effects* (RE) estimator. However, our objective is not to analyse RE because its assumptions are unlikely to be satisfied in practice. Also, we are not focusing on unbalanced datasets, whose discussion is postponed to future analysis while addressing the issue of attrition in panel data.

be $\beta_k = 1$, for $k = 1, \ldots, 4$, and $\beta_5 = 0$; $\theta = 0$ because errors are conditionally serially uncorrelated by assumption as in Stock and Watson (2008); the degree of heteroskedasticity assumes values of $\gamma = \{0, 2\}$, where $\gamma = 0$ stands for homoskedasticity and $\gamma \gg 1$ for severe heteroskedasticity. The scaling factor, $z(\gamma)$, is chosen such that the average variance of the error term is equal to one[11].

The contamination of random variables with good leverage points is completely random over the observations (i.e., *cell-isolated* anomalous cases). Good leverage points are obtained by randomly replacing 10% of the values[12] of x_1 with extreme observations drawn from a normal distribution with mean $\mu_{x_1} = 5$ and standard deviation $\sigma_{x_1} = 25$. Because x_1 is contaminated, then the variables generated from the former are directly affected by this source of contamination. The remaining random variables – x_3, x_4, x_5 – are either generated from the square or the product of x_1 and x_2 and, hence, follow a $\chi^2_{(v_1)}$ and a Gamma distribution, respectively.

The model is estimated including the set of aforementioned time-varying covariates and individual specific fixed effects, α_i. We estimate model (1.17) using fixed effects (FE) by applying the *within-group* (or time-demeaning) transformation to simulated data. Then, we estimate the time-demeaned regression specification using OLS[13]. As in Hansen (2007), the DGP for the simulations involves only random effects (RE) model because with (1.20) we assumed that the unobserved fixed effect is uncorrelated with the regressors. The model could be estimated more

[11]The error term u_{it} is intrinsically heteroskedastic but not on average due to the presence of the scaling factor $z(\gamma)$. The distribution of the random variable $W = \beta_0 + \sum_{j=1}^{J} \beta_j x_{it,j}$ and W^2 is provided in Appendix A.5. The algebraic derivation of the means and variances are shown.

[12]The degree of contamination could have been set to be even more or less severe according to the relevance the researcher attributes to the presence of extreme observations in the sample.

[13]We do not use the FGLS-FE to estimate the estimating Equation (1.17) for three main reasons. First, when the sample size is not sufficiently large there is an efficiency loss with respect to the FE-OLS estimator. In this analysis, we are interested in investigating the finite sample properties of the estimator, when N is not very large. Second, the FGLS-FE procedure requires to drop one of the time periods because the variance matrix is not invertible, leading to the reduction of the (already small) panel sample size (Cameron and Trivedi, 2005, ch.21.6, p.729). Third, FGLS-FE relies on the quality of the estimation of the variance and on the knowledge of the form of heteroskedasticity. However, the form of heteroskedasticity is always unknown from the data and the researcher has to make assumptions on the relationship between the variance of the disturbances and observables and unknown parameters (Cameron and Trivedi, 2005, Chapter 21, pp. 720-721, 729). This is unpractical in many areas of application and subjective to the researcher's guess. To overcome this limit, an objective criterion that has become a standard practice in applied works consists in using conventional robust standard errors due to software facilities (Verbeek, 2008).

efficiently with RE but FE models are commonly used in empirical studies with panel data[14].

Our Monte Carlo simulation involves 10,000 replications. The simulations are run for a combination of cross-sectional units $N = \{25, 50, 150, 500\}$ and time periods $T = \{2, 5, 10, 20\}$. Both cross-sectional units and time periods can be grouped as small ($N = \{25, 50\}$; $T = \{2, 5\}$), moderately small ($N = \{150\}$; $T = \{10\}$), and moderately large ($N = \{500\}$; $T = \{20\}$). The simulation is programmed in STATA16-MP and the main procedure is implemented in MATA.

1.8 Testing the Performance of HC Estimators

We examine the performance of each estimator in terms of proportional bias (PB), rejection probability (RP, or empirical size), adjusted power test, and root mean squared error (RMSE). Results are provided for a battery of estimators by a combination of panel units, time periods, and degree of heteroskedasticity, $\{N, T, \gamma\}$, where the number of units N varies in an interval from 25 to 500 units, time is fixed at $T = \{2, 5, 10, 20\}$, and the parameter that controls for the degree of heteroskedasticity is $\gamma \in \{0, 2\}$. This design is in accordance with the finite T assumption in the model as time periods are fixed while the number of observations increases.

Good leveraged data and heteroskedasticity make, as expected, test statistics calculated with conventional robust standard errors over-sized, upward biased, and with low power when the panel size $n < 2,500$. The proposed PHC6 mimics the behaviour of PHCjk in terms of PB, RP and power in all samples. PHC3 shows similar patterns but with different magnitudes.

[14] In future analysis we will re-assess the current version of the Monte Carlo simulation allowing α_i to be correlated with $\mathbf{x}_i$ to satisfy FE assumptions. Under this simulation design, $\widehat{\beta}$ estimated with FE remains consistent but is less efficient than $\widehat{\beta}$ estimated with RE.

1.8.1 Rejection Probability and Probability Bias

RP (i.e., the size of a test) in a Monte Carlo exercise with R runs is the frequency at which a rejection of the true null hypothesis occurs on average. A test statistics has a good size if rejects the null hypothesis approximately around the chosen $\alpha\%$ of the simulations, when the model is generated under the assumption that the null hypothesis is actually true.

The steps to obtain the empirical size in a two-sided single coefficient test are as follows. First, for each combination of $\{N,T\}$ and each simulation run $r = 1,\ldots,R$, compute the test statistics under the true null hypothesis,

$$T^0_{N,T}(\widehat{\beta}_{N,T,r}) = \frac{(\widehat{\beta}_{N,T,r} - \beta^0)}{\sqrt{\widehat{\text{AVar}(\widehat{\beta}_{N,T,r})}}} \overset{a}{\sim} t_{(df_r,\alpha/2)}.$$

Second, set the indicator $\mathbb{1}\{\cdot\}$ to turn on when the null hypothesis is rejected according to the rule

$$J^0_{N,T,r}(\widehat{\beta}) \equiv \mathbb{1}\left\{\left|T^0_{N,T}(\widehat{\beta}_{N,T,r})\right| > t_{(df_r,\alpha/2)}\right\},$$

where $t_{(df_r,\alpha/2)}$ is the critical value from a student-t distribution with df_r, degrees of freedom for a two-sided hypothesis test[15]. Third, count the total number of times a rejection has occurred and average it out by the number of replications R; the empirical size denotes the percentage of rejections in the Monte Carlo exercise as

$$\bar{J}^0_{N,T,r}(\widehat{\beta}) \equiv \frac{1}{R}\sum_{r=1}^{R} J^0_{N,T,r}(\widehat{\beta}) = \alpha_{test}.$$

For a two-sided test with q linear restrictions, the coverage probability is computed as follows. First, for each combination of $\{N,T\}$ and each simulation run $r = 1,\ldots,R$, compute the Wald

[15] With non-clustered inference $df_r = (NT-1)-(N+k-1)$ otherwise $df_r = N-1$.

statistics under the true null hypothesis, $H_0 : \mathbf{R}\boldsymbol{\beta} - \mathbf{r}^0 = \mathbf{0}$,

$$W^0_{N,T,r}(\widehat{\boldsymbol{\beta}}) = N(\mathbf{R}\widehat{\boldsymbol{\beta}}_{N,T,r} - \mathbf{r}^0)' \left\{ \mathbf{R}\widehat{\text{AVar}(\widehat{\boldsymbol{\beta}}_{N,T,r})}\mathbf{R}' \right\}^{-1} (\mathbf{R}\widehat{\boldsymbol{\beta}}_{N,T,r} - \mathbf{r}^0) \overset{a}{\sim} \chi^2(q),$$

where $\mathbf{R}$ is a $q \times K$ matrix with $q \leq K$, and $\mathbf{r}^0$ is a $q \times 1$ vector. Second, Mark as one every time a rejection occurs according to the rule

$$\tilde{J}^0_{N,T,r}(\widehat{\boldsymbol{\beta}}) \equiv \mathbb{1}\left\{ W^0_{N,T,r}(\widehat{\boldsymbol{\beta}}) > cv_{\chi^2(q)} \right\},$$

where $cv_{\chi^2(q)}$ is the critical value from a χ^2 distribution with q degrees of freedom for a two-sided hypothesis test[16]. Third, sum the cases when the null hypothesis has been rejected according to the above rule, and divide the number by the total number of simulation runs. The empirical size for a joint coefficient test is given by the percentage of rejections in the overall Monte Carlo as follows

$$\bar{\tilde{J}}^0_{N,T,r}(\widehat{\boldsymbol{\beta}}) \equiv R^{-1} \sum_{r=1}^{R} \tilde{J}^0_{N,T,r}(\widehat{\boldsymbol{\beta}}) = \alpha_{test}.$$

In the simulations, we test $H_0 : \beta_j = 1$ against $H_1 : \beta_j \neq 1$ for $j = 1$ while in a two-sided joint test we test $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 1$ against H_1 : at least one $\beta_j \neq 1$, for $j = 1, \ldots, 4$. The closer the rejection probability is to the nominal level of 5%, the better the estimator's performance in terms of empirical size (or type I error).

The proportional bias (PB) is a measure of the bias of the estimator of the variance-covariance matrix computed as $PB = 1 - SE(\hat{\beta}_j)/SD(\hat{\beta}_j)$, where *SE* stands for standard error and *SD* for standard deviation. Positive (negative) values of PB indicate by how much the standard error obtained using one of the four formulae presented above underestimates (overestimates) the

[16] Alternatively, the F statistic can be computed from the Wald test statistics as $F^0_{N,T,r}(\widehat{\boldsymbol{\beta}}) = W^0_{N,T,r}(\widehat{\boldsymbol{\beta}})/q \overset{a}{\sim} F_\alpha(q, df_r)$ under the true null hypothesis, where q are the number of restrictions and degrees of freedom at the numerator, and df_r are the residual degrees of freedom or degrees of freedom at the denominator.

"true" standard error.

In this section, we comment on the performance of each estimator taking into account its ability to reject the true null hypothesis at 5% significance level along with its accuracy. Tables B.1 and B.2 report the results of the Monte Carlo simulations respectively, with and without heteroskedasticity. Each table compares the PB, RP and RMSE[17] of four alternative formulae of the variance-covariance matrix (i.e., PHC0, PHC3, PHC6 and PHCjk). Results are grouped by different combinations of sample size N and time length T. Figures refer to the slope parameter β_1, which is associated with the contaminated variable $x_{it,1}$. The t-test statistics are at 5%-level.

Under heteroskedasticity, PHC0 standard errors considerably underestimate the "true" variance (positive PB) on average by at least 30% when $n \leq 2,500$. PHC6 mimics the behaviour of PHCjk in small and large samples, overestimating the true variance (negative PB: min= 1.2% and max = 12.3%) for $n \leq 300$ and slightly underestimating the true variance (positive PB: min= 4.9% and max = 10.6%) in the other cases. For $N = \{25, 50\}$ and all T the PB of PHCjk is larger in absolute value than PB of PHC6 if the bias is positive, and smaller otherwise. From $N \geq 150$ PHCjk and PHC6 produce the same bias but PHC3 produces a smaller bias in absolute value when the estimators over-estimate the variance.

Test statistics of PHC0 are largely over-sized (RP above 0.05) when $N = \{25, 50\}$ and all T but approach the true α%-size when $n \geq 5,000$, despite the high positive PB. The most conservative estimators always under-reject the null hypothesis (RP below 0.05), and as the cross-sectional size increases (fixing the time dimension) the RP gradually converges to 5% but their test statistics still remain slightly under-sized. However, looking at the (positive/negative) distance from 0.05 PHC0 turns out to be more over-sized then he other estimators when $n \leq 750$.

In general, a smaller PB in absolute value (signaling a good approximation of the "true" variance) does not automatically imply that the empirical size is the closest to the actual nominal

[17] Results for the RMSE are commented in Section 1.8.2.

significance level. The "true" standard errors remain under-estimated (over-estimated) if the bias is positive, despite producing test statistics that reject the null hypothesis with much precision.

Under homoskedasticity, PHC0 always underestimates the true variance (especially for $n \leq 1,500$). The PB reduces as the panel size increases but only when $n = 10,000$ it drops considerably. The other PHC estimators tend to over-estimate the true variance (negative PB) but the magnitudes are smaller in absolute value than the figures of PHC0. PHC6 has similar bias to PHCjk while PHC3 is slightly more biased. Test-statistics of PHC0 are over-sized (large RP) for $n \leq 1,500$ but show a convergence pattern to 5% as the sample size increases. The test size of PHC6 and PHCjk is always closest to the true α-size followed by PHC3.

Tables B.3 and B.4 report the Wald test statistics and RP from the joint coefficient test for the slope coefficients different from zero (i.e., β_i for $i = 1, \ldots, 4$) under heteroskedasticity and homoskedasticity, respectively. Results for different combinations of $\{N, T\}$ are displayed. The nominal level of significance is set at $\alpha = 0.05$. The closer the value of the rejection rate of the test statistic is to $\alpha = 0.05$, the better the estimator's performance in terms of empirical size.

Under heteroskedasticity and good leveraged data points, the RPs of the four estimators slowly converge to 5% as the sample size increases with exception of PHC6. PHC6 is outperformed by PHC0 in terms of RP for $n \geq 2,500$. Despite the upward distortion of all test statistics for $N \leq 500$ and all T, PHC0 raw sizes are the largest in magnitude among the four estimators in very small cross-sectional samples. The Wald statistics of the other two conservative estimators are the lowest in magnitudes for $n \geq 300$. Similar patterns are observed under the assumption of homoskedasticity. PHC0 performs as well as the two conservative estimators only for large N ($N \geq 150$ fixing T).

1.8.2 RMSE Assessment

An additional evaluation on the quality of the four estimators is done in terms of the RMSE. For each estimator of the variance, the RMSE is computed as the square root of the average deviation of the standard error from the standard deviation of the estimated coefficient of β_j. In formulae,

$$\text{RMSE}_j^s = \frac{1}{R}\sum_{r=1}^{R}\sqrt{(\hat{\sigma}^s(\hat{\beta}_j)_r - \sigma(\hat{\beta}_j)_r)^2} \tag{1.23}$$

where $\hat{\sigma}^s(\hat{\beta}_j)_r$ is the standard error of $\hat{\beta}_j$ in the rth run of the simulation computed using one of the HC formulae, and $\sigma(\hat{\beta}_j)_r$ is the standard deviation of the estimated coefficient β_j. A good quality estimator has its RMSE close to zero. Because the RMSE and PB are constructed from the same quantities, $\hat{\sigma}^s(\hat{\beta}_j)$ and $\sigma(\hat{\beta}_j)$, they are linked one to the other. The larger the proportional bias in absolute value, the larger the RMSE of the estimator is in magnitude.

Results are presented in Table B.1 and B.2 for different combinations of cross-sectional units and time length, and under different degrees of heteroskedacity. Under heteroskedasticity, the RMSE of PHC0 estimator is much higher than those of the other three estimators for all combinations of N and T. The RMSE of the three conservative estimators gradually converges to zero in large samples, displaying similar values in small samples. Under homoskedasticity, the RMSE of all estimators are always very close to zero for different combinations of panel sample size. The only exception is for $n \leq 100$ when PHC6 has the smallest RMSE.

1.8.3 Adjusted Power Test

The power of the test is the average frequency at which the false null hypothesis is rejected in a simulation. In a two-sided single coefficient test, the adjusted power for the false null hypothesis is obtained through the steps below. First, for each combination of $\{N,T\}$ and for each simulation run $r = 1,\ldots,R$, compute the test statistics under the false null hypothesis as

$$T^1_{N,T}(\widehat{\beta}_{N,T,r}) = \frac{(\widehat{\beta}_{N,T,r} - \beta^1)}{\sqrt{\widehat{\text{AVar}(\widehat{\beta}_{N,T,r})}}} \overset{a}{\sim} t_{(df_r,\alpha/2)}.$$

Second, the indicator $\mathbb{1}\{\cdot\}$ turns on every time that the rejection rule holds

$$J^1_{N,T,r}(\widehat{\boldsymbol{\beta}}) \equiv \mathbb{1}\{T^1_{N,T}(\widehat{\beta}_{N,T,r}) < \mathbf{t}^0_{\alpha/2} \text{ or } T^1_{N,T}(\widehat{\beta}_{N,T,r}) > \mathbf{t}^0_{1-\alpha/2}\},$$

where $\mathbf{t}^0_{\alpha/2}$ and $\mathbf{t}^0_{1-\alpha/2}$ are values lying respectively at the $(\alpha/2)^{th}$ and $(1-\alpha/2)^{th}$ percentiles of $T^0_{N,T}(\widehat{\beta}_{N,T,r})$, and used as critical values[18]. The empirical critical values differ due to the asymmetric distribution of the test statistics. Third, count the total number of rejections in the simulation and divide by the number of runs; the adjusted power of a test is

$$\bar{J}^1_{N,T,r}(\widehat{\boldsymbol{\beta}}) \equiv R^{-1}\sum_{r=1}^{R} J^1_{N,T,r}(\widehat{\boldsymbol{\beta}}) = 1 - \theta_{test}.$$

Similarly, for a two-sided test with q linear restrictions the adjusted power of a test is conducted as follows. First, for each combination of $\{N,T\}$ and for each simulation run $r = 1,\ldots,R$, compute the Wald statistics under the true null hypothesis, $H_0 : \mathbf{R}\boldsymbol{\beta} - \mathbf{r}^1 = \mathbf{0}$,

$$W^1_{N,T,r}(\widehat{\boldsymbol{\beta}}) = N(\mathbf{R}\widehat{\boldsymbol{\beta}}_{N,T,r} - \mathbf{r}^1)'\Big\{\mathbf{R}\widehat{\text{AVar}(\widehat{\boldsymbol{\beta}}_{N,T,r})}\mathbf{R}'\Big\}^{-1}(\mathbf{R}\widehat{\boldsymbol{\beta}}_{N,T,r} - \mathbf{r}^1) \overset{a}{\sim} \chi^2(q),$$

where $\mathbf{r}^1$ is a $q \times 1$ vector. Second, define the F statistics $F^1_{N,T,r}(\widehat{\boldsymbol{\beta}}) = W^1_{N,T,r}(\widehat{\boldsymbol{\beta}})/q$ under the false null hypothesis for replication run r, and sample combination $\{N,T\}$. The rejection rule is defined as

$$\tilde{J}^1_{N,T,r}(\widehat{\boldsymbol{\beta}}) \equiv \mathbb{1}\{F^1_{N,T,r}(\widehat{\boldsymbol{\beta}}) > F^0_{\alpha}\},$$

[18] We cannot use conventional critical values from the t-distribution because size-unadjusted power curves make any comparison between estimators meaningless.

where F^0_α is the value lying at the α^{th} quantile of distribution of $F^0_{N,T,r}(\widehat{\beta})$ derived under the true null hypothesis, and used as empirical critical in the rejection rule. Third, the percentage of rejections that occur in the Monte Carlo exercise is the adjusted power of a test,

$$\bar{\tilde{J}}^1_{N,T,r}(\widehat{\beta}) \equiv R^{-1}\sum_{r=1}^{R}\tilde{J}^1_{N,T,r}(\widehat{\beta}) = 1-\theta_{test}.$$

In the simulations, we test $H_0 : \beta_j = 1$ against $H_1 : \beta_j \neq 001$ for $j = \{1,2\}$ for two-sided single coefficient tests, where β^1 is a value taken from a narrow interval around the true β_j. For two-sided joint tests we test $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 1$ against H_1 : at least one $\beta_j \neq 1$, for $j = 1,\ldots,4$.

Figures C.2 and C.3 plot size-adjusted power curves of a battery of HC estimators for different panel sample sizes and degree of heteroskedasticity for β_1. The vertices of all power curves correspond to the nominal size of the test statistics, $\alpha = 0.05$. It is common practice to adjust the power for the empirical size because the empirical distributions of test statistics may depend on the nature of the specific regressor and, therefore, any comparison across estimators turns out to be meaningless without size-adjustment. Precisely, in the absence of any size-adjustment the most liberal estimator would tend to have greater power than the most conservative estimator because the former is more likely to over-reject the null hypothesis in favour of the alternative, while the opposite is true for the latter. Unlike the test size, simulation results for the test power do not differ considerably in terms of the overall pattern, but they do in terms of magnitudes.

Under heteroskedasticity, simulation results show that PHC0 does not have as good power performance as PHC3, PHCjk and PHC6 in small samples ($N = \{25,50\}$ and especially with $T = 2$). In fact, its rejection probabilities at a given parameter value are lower than those of the other three estimators. Fixing T and letting N change, the power performance of PHC0 does not improve. Rejection probabilities remain the lowest and slowly converge to one, even when the distance from the true value of β increases. Conversely, we do not observe such a remarkable loss in power when we let T increase and fixing N as the difference with other estimators in the rejection probabilities at a given parameter value becomes negligible or vanishes completely.

Under the assumption of homoskedasticity, PHC0 has better power than PHC3, PHCjk, and PHC6 with $N = 25$ for all T. This result is in stark contrast with PHC0 poor test size (i.e., RP) described above due to the usual trade-off between type I and type II error. When $T = 2$, all power curves show a lower convergence to one.

Figures C.4 and C.5 show the adjusted power curves for the joint coefficient test. From the graphs we observe that all power curves are well-behaved under homoskedasticity with rejection rates approaching one quite rapidly as the tested parameter values depart from the true value, and with the increase in the sample size. This cannot be said under heteroskedasticity and, especially, when the panel sample size is small (small N and small T) because test statistics of all estimators have low rejection power, especially PHC0 test statistics when $N = \{25, 50\}$ and $T = \{2, 5\}$.

Overall, the four estimators have similar asymptotic behaviour with or without heteroskedasticity. This can be explained by the sensitivity of the test of hypothesis to sample size. In fact, as the sample size increases the probability of rejecting the false null hypothesis (i.e., the power of the test) increases as well, by construction. The opposite happens to the size of a test instead.

1.9 Conclusion

In this chapter, we investigated the effects of the simultaneous presence of a small sample size, heteroskedasticity, and good leveraged data on the validity of conventional statistical inference in linear panel data models with fixed effects. We documented their detrimental effects on the statistical inference calculated with robust standard errors. More conservative estimators of the sampling variance produce test statistics that have unbiased empirical sizes and higher power under these circumstance.

We formalised a panel version of MacKinnon and White's (1985) and Davidson et al.'s (1993) estimators, and proposed a new hybrid estimator, *PHC*6. We derived the finite sample properties and the asymptotic distributions of the discussed HC estimators. With MC simulations we compared the performances of four types of standard errors, computed with Arellano's (1987) and three types of jackknife-like formulae, in terms of empirical size and power. We documented the downward bias of conventional robust standard errors under specific circumstances, suggesting alternatives to obtain more reliable statistical inference.

The main findings can be summarised as follows. Under heteroskedasticity, more conservative standard errors should be used in the presence of leverage points because their test statistics possess a low proportional bias, small size distortions, and have higher power. Conversely, conventional standard errors and the proposed formula, PHC6, should be preferred with homoskedasticity because the other conservative estimators excessively under-reject the true null hypothesis. Under homoskedasticity cluster-robust formulae should always be used. A similar result was found in MacKinnon and White (1985) and Long and Ervin (2000) for cross-sectional models. The cross-sectional dimension matters for the finite sample properties of the estimators but not the size of N relative to T. However, conventional cluster-robust standard errors remain upward biased even when their empirical size is correct, and even in larger samples.

Chapter 2

Influence Analysis in Panel Data Models

2.1 Introduction

Econometric techniques based on least squares (LS) are, by construction, extremely sensitive to the presence of three different types of "anomalous" points. First, *vertical outliers* (henceforth, VO) do not follow the general pattern of the rest of the cloud of data. Second, *good leverage* points (GL points) exhibit extreme values in the covariates but still lie on the regression line. Third, *bad leverage* points (BL points) are a combination of the two. These points are known to exert a disproportionate influence on the estimates of linear regression models leading to biases in the regression estimates – on estimated coefficients and/or standard errors (Donald and Maddala, 1993; Bramati and Croux, 2007; Verardi and Croux, 2009), as shown in Figure C.6.

Appropriate tools for the detection of such critical observations serve to identify them and assess their overall influence on the estimates. Because the case-wise deletion may substantially alter the features of LS analysis, it is a common practice to examine the outlyingness of deleted cases on the basis of the distance of the observation from the mean of the data relative to its dispersion (Cook, 1979; Rousseeuw, 1991; Poon and Poon, 2001). Following this idea, a popular measure for the detection of influential cases is the Cook's (1979) distance. A main limitation of this metric is that it fails to detect the simultaneous effect of multiple atypical cases (Atkinson, 1985; Chatterjee and Hadi, 1988; Rousseeuw and Van Zomeren, 1990). This phenomenon is known as the *masking effect*. Conversely, measures based on a local approach can handle multiple atypical cases at the same time (Lawrance, 1995; Poon and Poon, 2001).

This chapter formalises statistical measures for quantifing the degree of leverage and outlyingness of units in a panel data framework. We develop a method to visually detect the type of anomaly and quantify its joint and conditional effects, when the *within group* transformation is applied to the original data points.

We first formalise the notion of the average individual leverage and average normalised residuals used for unit-wise leverage-residual plots[1]. We then propose two diagnostic measures for panel data – based on Lawrance's (1995) joint and conditional measures – for the evaluation of the joint and conditional influence of pairs of units on the LS estimates. Then, we use tools from the network analysis to show the overall and bilateral influence of highly influential units. We finally apply our influence analysis to four macroeconomic country-level studies published articles in the American Economic Review. We replicate their main regression results using alternative formulae of standard errors.

We observe that joint and conditional diagnostic measures are, by construction, better in the detection of leveraged units. In fact, BL and GL units possess the highest total joint and conditional influence and contribute to enhancing and masking the effects of even fairly influential units. Conversely, VO do not contribute in exerting large total influence and, hence, in affecting a wider cloud of units. Whenever we detect GL points in the replicated studies, the statistical inference conducted with cluster-robust standard errors is over-inflated, leading to invalid conclusions.

The novelty of this research consists in: (i) visually inspecting the overall contribution

[1] At the moment of writing, the STATA command `lvr2plot` is available for a cross-sectional evaluation of each case in the sample. We write its panel counterpart `xtlvr2plot` based on the quantities defined in Section 2.4.1.

of the most influential units in the sample with a network graphical representation; and (ii) using a "unit-wise" approach instead of a "case-wise" evaluation in the deletion procedure[2].

The existing literature on diagnostic measures for the detection of anomalous data points is somewhat limited for both cross-sectional and panel data models. Cook's (1979) distance is still the most popular measure for the detection of individually influential cases. Following Cook (1979), authors like Banerjee and Frees (1997), Martín and Pardo (2009), Pinho et al. (2015) and Martín (2015) proposed similar diagnostic measures for other econometric models. However, the limit of Cook-like measures in detecting the reciprocal influence of groups of observations is widely documented such that the effect of other cases is hidden (Atkinson, 1985; Chatterjee and Hadi, 1988; Rousseeuw and Van Zomeren, 1990; Lawrance, 1995; Poon and Poon, 2001; Kim, 2017). This limitation was overcome by Lawrance (1995) who conceptualised two measures for quantifying the joint and conditional effects of pairs of observations on the estimated parameters with cross-sectional data. But these two diagnostic measures have never been implemented in practice[3]. We contribute to this literature by quantifying statistical metrics for a unit-wise evaluation of the type of anomaly, and formalising two diagnostic measures for the joint and conditional influence in panel data models with their statistical distribution. In this framework, Cook's (1979) distance for panel data is as a special case of the measure for joint influence, when unit i coincides with j.

The rest of the chapter is structured as follows. We start with the definition of the three types of anomalous points in a panel data framework in Section 2.2. Then, we introduce the LS estimators and their distributional results in Section 2.3. We then present diagnostic measures for panel data and their exact and asymptotic distributions in Section 2.4. We proceed with the visual

[2] In the "unit-wise" approach, the full history of the unit is dropped. This approach is more appropriate for the panel data framework for two reasons. First, a single influential case may not compromise the whole time series. Second, atypical observations may be carried over the full history of a unit so that the time dimension may contribute to exacerbate the effect of groups of influential observations on the estimates of interest (Bramati and Croux, 2007).

[3] At the time of writing, there is no STATA command to calculate these measures with cross-sectional data.

inspection of the points using the aforementioned tools in Section 2.5. Empirical examples are provided in Section 2.6. Section 3.5 concludes.

2.2 Influential Units in Longitudinal Data

Longitudinal data are extremely likely to contain units with values of the dependent and/or independent variables that follow a different pattern from the main cloud of the data in the response and/or covariate spaces. Their influence is assessed in terms of how their deletion affects regression estimates. According to their features, these points can be classified as VO, GL or BL points.

Vertical outliers are atypical points in the response-factor space with large squared normalised residual that follow the opposite trend of the cloud of data points, lying far from the regression line as if they were generated from a different process (Chatterjee and Hadi, 1986; Greene, 2012, p.141). Their presence alters the estimated LS intercept (Verardi and Croux, 2009), undermining the accuracy of the estimator but not its precision. Outliers cannot always be detected from the values of the LS residuals when the effect of an outlying observation is *masked* by a nearby observation that exerts a greater influence in attracting the regression fit (Rousseeuw and Van Zomeren, 1990). Removing bad realisations from the sample may not necessarily be the most accurate, albeit common, approach to follow without understanding the source of anomaly – e.g., a recording error, a different data generating process, or an incorrect specification of the model – and with the risk of leaving the undetected cases (Rousseeuw, 1991). In fact, it is not unlikely that in the presence of multiple outliers, an influential case is masked by another case in the sample. This is an instance of the masking effect.

Leverage points are observations that appear isolated from the rest of the data points but follow the same trend of the rest of the data (Chatterjee and Hadi, 1986; Greene, 2012, p.140). Leverage points can be distinguished in “good” or “bad” anomalous observations. The former exhibit unusually extreme values in the covariates and lie on the predicted regression line enhancing the precision of the regression fit; the latter possess extreme values in both input and response direction, lie far from the plane where the bulk of data points are and are not fitted by the regression model (Rousseeuw, 1991; Bramati and Croux, 2007). While "bad" leveraged data adversely affect the estimated LS coefficients (both the intercept and the slope), "good" leverage points add variability in the sample allowing for a better fit of the data (in Figure C.6 the dash line overlaps the red dotted line) at the cost of a deteriorated statistical inference (Silva, 2001; Verardi and Croux, 2009). GL points exert higher influence on the statistical significance of a coefficient inflating the strength of the regression relationship. Then, conventional Heteroskedasticity-Consistent (HC) standard errors become systematically downward biased, especially in small cross-sectional samples (MacKinnon and White, 1985; Chesher and Jewitt, 1987).

A visual inspection of the data is unlikely to be trivial in a k-dimensional space of covariates (Rousseeuw, 1991; Bramati and Croux, 2007). Leverage points are frequently found with observational data because the values of covariates are not fixed, reflecting the peculiarities of the dataset (Rousseeuw and Van Zomeren, 1990; Silva, 2001).

Unlike the cross-sectional setting, Bramati and Croux (2007) observe that with longitudinal data "anomalous" observations may appear either as isolated cases of different units (*cell-concentrated* points) or concentrated in the time-series of the same units (*block-concentrated* points). A univariate example that illustrates Rousseeuw's (1991) and Bramati and Croux's (2007) classifications is provided in Figure C.6, which depicts three types of "anomalous" observations that may appear in isolated cells or in blocks. In this explicative example, we use a simulated data set based on Bramati and Croux's (2007) design. In all scatter plots, unit 1 identifies "good" leverage points, unit 2 vertical outliers, and unit 3 "bad" leverage points. From left to right and from top to bottom, each pair relates the dependent and independent variable in the original data set, after the within-group transformation, and after the first-differencing. The main difference in the detection of anomalous units between cross-sectional and panel data consists in the type of data transformation used to estimate the longitudinal model. That is, both *within-group* and *first-differencing* transformations alter the original value of the time-series: the former is mean sensitive to extreme observations as data are centered around each time series whereas the latter is sensitive to changes in two consecutive years. Therefore, atypical cases should be detected afterwards if not estimating a Pooled OLS (POLS) model.

From the examples provided in Figure C.6, we can observe that the *within-group* transformation amplifies the anomaly arising from atypical blocks in the original sample because the mean is heavily affected by extreme values, by construction. Consequently, the uncontaminated cases of the time-series become extreme values with opposite sign after the data transformation. Clearly, the *withing-group* transformation exacerbates the problems caused by block anomalous cells because the uncontaminated cases of the times series become vertical outliers themselves with opposite sign with respect to the original contaminated block. However, this transformation attenuates the outlyingness of the original data because it is scaled-down, by construction. The *within-group* transformation seems to improve the accuracy of the fit with good leverage points, despite a deterioration in the statistical inference.

With *first-differencing*, anomalous blocks are reduced to anomalous isolated cells after the transformation due to the peculiar structure of the data design. In fact, in this sample design "anomalous" block-concentrated units are generated by contaminating half of the time-series ($T = 5$) with extreme subsequent cases. *First-differencing* can be a solution to minimise the impact of subsequent block-concentrated atypical observations because, by construction, only the difference between an atypical and normal value will generate an anomalous case at time t. However, this data transformation generates greater dispersion with not subsequent isolated cells. Without any panel data transformation, POLS fits a cross-sectional model using original data, whose parameter estimates are affected by raw individual cells or concentrated blocks of the affected units. This results in the worse data fit in the presence of outlying observations.

Overall, it is worth to remark that each data transformation may reduce or magnify the effects of the presence of anomalous cases on the estimated parameters or the model fit. This depends on the structure of the data – whether they are concentrated in part of the time series or isolated cells – and on the properties of each transformation.

2.3 The Estimators

Consider the static linear panel regression model with one-way error component

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\alpha}_i + \mathbf{u}_i, \text{ for all } i = 1, \ldots, N, \tag{2.1}$$

where $\mathbf{y}_i$ is $T \times 1$ vector of outcomes; $\mathbf{X}_i$ is a $T \times k$ matrix of time-varying regressors; $\boldsymbol{\alpha}_i = \alpha_i\boldsymbol{\iota}$ is a $T \times 1$ vector of individual fixed effects, and $\boldsymbol{\iota}$ is a vector of ones of order T; and $\mathbf{u}_i$ is a $T \times 1$ vector of one-way error component.

After applying the *within-group* transformation to model (2.1), the estimating equation becomes

$$\widetilde{\mathbf{y}}_i = \widetilde{\mathbf{X}}_i\boldsymbol{\beta} + \widetilde{\mathbf{u}}_i, \text{ for all } i = 1, \ldots, N, \tag{2.2}$$

where $\widetilde{\mathbf{y}}_i = (\mathbf{I}_T - T^{-1}\boldsymbol{\iota}\boldsymbol{\iota}')\mathbf{y}_i$ is $T \times 1$; $\widetilde{\mathbf{X}}_i = (\mathbf{I}_T - T^{-1}\boldsymbol{\iota}\boldsymbol{\iota}')\mathbf{X}_i$ is $T \times k$; and $\widetilde{\mathbf{u}}_i = (\mathbf{I}_T - T^{-1}\boldsymbol{\iota}\boldsymbol{\iota}')\mathbf{u}_i$ is $T \times 1$. Note that $(\mathbf{I}_T - T^{-1}\boldsymbol{\iota}\boldsymbol{\iota}')\boldsymbol{\alpha}_i = \mathbf{0}$ as $T^{-1}\boldsymbol{\iota}\boldsymbol{\iota}'\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_i$.

The OLS estimator of (2.2) is the *within-group* estimator of the true population parameter with formula $\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N} \widetilde{\mathbf{X}}_i'\widetilde{\mathbf{X}}_i\right)^{-1} \sum_{i=1}^{N} \widetilde{\mathbf{X}}_i'\widetilde{\mathbf{y}}_i$, which is unbiased and asymptotically normally distributed under conventional assumptions – i.e., linearity, exogeneity, full rank condition, no serial correlation (see ASM.1-ASM.5 in Chapter 1).

Removing one unit at a time from the sample gives a general measure of the influence exerted by that unit on the parameter estimates or in the fit of the model. The *within-group* estimator without the whole history of unit i is the Leave-One-Out (L1O) estimator

$$\widehat{\boldsymbol{\beta}}_{(i)} = \widehat{\boldsymbol{\beta}} - \left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i \tag{2.3}$$

where $\mathbf{M}_i^{-1} = (\mathbf{I}_T - \mathbf{H}_i)^{-1}$ with $\mathbf{H}_i = \widetilde{\mathbf{X}}_i(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}_i'$, and $\widehat{\mathbf{u}}_i = \widetilde{\mathbf{y}}_i - \widetilde{\mathbf{X}}_i\widehat{\boldsymbol{\beta}}$ is the residual term. Formula (2.3) is extensively derived in Appendix A.1 following Banerjee and Frees (1997). The distribution of $\widehat{\boldsymbol{\beta}}_{(i)}$ is derived in Appendix A.7.

The deletion of pairs of units $\{i, j\}$ at a time from the sample captures the influence exerted by that pair on the parameter estimates or in the fit of the model. The *within-group* estimator without the full history of units i and j is the Leave-Two-Out (L2O) estimator

$$\widehat{\boldsymbol{\beta}}_{(i,j)} = \widehat{\boldsymbol{\beta}}_{(i)} - \left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\left(\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\mathbf{H}_{ij} + \widetilde{\mathbf{X}}_j'\right)\left(\mathbf{M}_j - \mathbf{H}_{ij}'\mathbf{M}_i^{-1}\mathbf{H}_{ij}\right)^{-1}\left(\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i + \widehat{\mathbf{u}}_j\right) \tag{2.4}$$

where $\mathbf{M}_j = \mathbf{I}_j - \mathbf{H}_j$ with $\mathbf{H}_{ij} = \widetilde{\mathbf{X}}_i(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}_j'$, and $\mathbf{H}_j = \widetilde{\mathbf{X}}_j(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}_j'$. Note that $\mathbf{H}_{ji} = \mathbf{H}_{ij}'$. Formula (2.4) is derived in Appendix A.6, and the distribution of $\widehat{\boldsymbol{\beta}}_{(i,j)}$ is presented in Appendix A.7.

2.3.1 Distributional Results in Panel Data Models

We first establish the asymptotic properties of $\widehat{\boldsymbol{\beta}}_{(i)}$ and show its equivalence to $\widehat{\boldsymbol{\beta}}$, implying that the two estimators share the same asymptotic distribution, where $\widehat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}$ with asymptotic distribution $\sqrt{N}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{S}_{XX}^{-1}\mathbf{V}\mathbf{S}_{XX}^{-1}\right)$ as N goes to infinity and T is fixed.

Applying the *Triangle Inequality* on Equation (2.3),

$$\begin{aligned}\|(\widehat{\boldsymbol{\beta}}_{(i)}-\boldsymbol{\beta})\| &\leq \|(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\| \\ &+\left\|\left(\frac{1}{N}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\right\|\frac{1}{N}\|\widetilde{\mathbf{X}}_i\|\|(\mathbf{I}_T-\mathbf{H}_i)^{-1}\|\left(\|\widetilde{\mathbf{u}}_i\|+\|\widetilde{\mathbf{X}}_i\|\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|\right) \\ &= o_p(1)\end{aligned} \tag{2.5}$$

where $\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|$ is $o_p(1)$; $\left(N^{-1}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}=\mathbf{S}_{XX}^{-1}+o_p(1)$ by ASM.3, *WLLN* and *Slutsky's theorem*; $\widetilde{\mathbf{X}}_i$ and $\widetilde{\mathbf{u}}_i$ are $O_p(1)$ because random variables with finite moments by ASM.4 and, therefore, $N^{-1}\|\widetilde{\mathbf{X}}_i\|=O_p(N^{-1})$; and $\|(\mathbf{I}_T-\mathbf{H}_i)^{-1}\|\leq\sqrt{T}+N^{-1}\|\widetilde{\mathbf{X}}_i\|^2\left\|\left(N^{-1}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}-N^{-1}\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{X}}_i\right)^{-1}\right\|$ is O(1) because the first term on the right-hand-side, $\sqrt{T}$, is O(1) without a remainder term, and the second component is bounded above by $o_p(1)$ random variable. Then, it follows that

$$\widehat{\boldsymbol{\beta}}_{(i)}=\boldsymbol{\beta}+o_p(1). \tag{2.6}$$

We now show that the estimators $\widehat{\boldsymbol{\beta}}_{(i)}$ and $\widehat{\boldsymbol{\beta}}$ have the same asymptotic distribution. Using the *Reverse Triangle Inequality*, we obtain

$$\begin{aligned}&\|\sqrt{N}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})-\sqrt{N}(\widehat{\boldsymbol{\beta}}_{(i)}-\boldsymbol{\beta})\| \\ &\leq\left\|\left(\frac{1}{N}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\right\|\frac{1}{\sqrt{N}}\left\|\widetilde{\mathbf{X}}_i\right\|\|(\mathbf{I}_T-\mathbf{H}_i)^{-1}\|\left(\|\widetilde{\mathbf{u}}_i\|+\|\widetilde{\mathbf{X}}_i\|\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|\right) \\ &= o_p(1)\end{aligned} \tag{2.7}$$

where the first component is $\left(\mathbf{S}_{XX}^{-1}+o_p(1)\right)$ as $N^{-1}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\xrightarrow{p}\mathbb{E}(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})=\mathbf{S}_{XX}>0$ by the *Central Limit Theorem*; the second term is ,$O_p(N^{1/r-1/2})=O_p(1)$ for $r\geq 2$ under ASM.4.i; the third component is $O(1)$; and the last quantity in parenthesis is $O_p(1)$. We can conclude that $\widehat{\boldsymbol{\beta}}_{(i)}$ and $\widehat{\boldsymbol{\beta}}$ share the has identical limiting distribution in the sense that

$$\sqrt{N}\big(\widehat{\boldsymbol{\beta}}_{(i)}-\boldsymbol{\beta}\big)=\sqrt{N}\big(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\big)+o_p(1). \tag{2.8}$$

The distance between $\widehat{\boldsymbol{\beta}}_{(i)}$ and $\widehat{\boldsymbol{\beta}}$ vanishes as the cross-sectional size grows and, hence, the influence exerted by unit i is null. Removing one unit does not have an impact on the estimates of the true value of the parameter as the cross-sectional units increase to infinity while keeping time fixed. The asymptotic variance of $\widehat{\boldsymbol{\beta}}_{(i)}$ converges to the variance of $\widehat{\boldsymbol{\beta}}$ as

$$\text{Avar}\Big(\sqrt{N}\big(\widehat{\boldsymbol{\beta}}_{(i)}-\boldsymbol{\beta}\big)\Big)=N\text{Var}\big(\widehat{\boldsymbol{\beta}}_{(i)}|\widetilde{\mathbf{X}}_i\big)\equiv\mathbf{V}_{\widehat{\boldsymbol{\beta}}}\quad\text{as } N\to\infty \text{ and } T \text{ fixed} \tag{2.9}$$

where $\mathbf{V}_{\widehat{\boldsymbol{\beta}}}=\mathbf{S}_{XX}^{-1}\boldsymbol{\Sigma}\mathbf{S}_{XX}^{-1}$ and

$$\begin{aligned}\text{Var}\big(\widehat{\boldsymbol{\beta}}_{(i)}|\widetilde{\mathbf{X}}_i\big)&=\mathbb{E}\Big\{\big(\widehat{\boldsymbol{\beta}}_{(i)}-\boldsymbol{\beta}\big)\big(\widehat{\boldsymbol{\beta}}_{(i)}-\boldsymbol{\beta}\big)'|\widetilde{\mathbf{X}}_i\Big\} &(2.10)\\ &=\text{Var}\big(\widehat{\beta}|\widetilde{\mathbf{X}}\big)+\text{Var}\big(\widehat{\beta}|\widetilde{\mathbf{X}}\big)\text{B}(\widetilde{\mathbf{X}}_i)'+\text{B}(\widetilde{\mathbf{X}}_i)\text{Var}\big(\widehat{\beta}|\widetilde{\mathbf{X}}\big) &(2.11)\\ &+\text{A}(\widetilde{\mathbf{X}}_i)\mathbf{M}_i^{-1}\boldsymbol{\Sigma}_i\mathbf{M}_i^{-1}\text{A}(\widetilde{\mathbf{X}}_i)'+\text{B}(\widetilde{\mathbf{X}}_i)\text{Var}\big(\widehat{\beta}|\widetilde{\mathbf{X}}\big)\text{B}(\widetilde{\mathbf{X}}_i)' &(2.12)\end{aligned}$$

where $\text{A}_N(\widetilde{\mathbf{X}}_i)\equiv\big(N^{-1}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\big)^{-1}N^{-1}\widetilde{\mathbf{X}}_i'$ and $\text{B}_N(\widetilde{\mathbf{X}}_i)=\text{A}_N(\widetilde{\mathbf{X}}_i)\mathbf{M}_i^{-1}\widetilde{\mathbf{X}}_i$. Under the model assumptions and using THM 6.6 in Hansen (2019, p.182), the matrices $\text{A}_N(\widetilde{\mathbf{X}}_i)\xrightarrow{p}\mathbf{0}$ and $\text{B}_N(\widetilde{\mathbf{X}}_i)\xrightarrow{p}\mathbf{0}$ as $N\to\infty$ and T fixed. Full proof is provided in Appendix A.7.

We follow a similar procedure to derive the distribution of $\widehat{\boldsymbol{\beta}}_{(i,j)}$. Using *Triangle In-*

equality,

$$\|\widehat{\boldsymbol{\beta}}_{(i,j)}-\boldsymbol{\beta}\| \leq \|\widehat{\boldsymbol{\beta}}_{(i)}-\boldsymbol{\beta}\| \tag{2.13}$$

$$+\left\|\left(\frac{1}{N}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\right\| \tag{2.14}$$

$$\left(\frac{1}{N}\|\widetilde{\mathbf{X}}_i\|\|\mathbf{M}_i^{-1}\|\|\mathbf{H}_{ij}\|+\frac{1}{N}\|\widetilde{\mathbf{X}}_j\|\right) \tag{2.15}$$

$$\left\|\left(\mathbf{M}_j-\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\mathbf{H}_{ij}\right)^{-1}\right\| \tag{2.16}$$

$$\left(\|\mathbf{H}_{ij}\|\|\mathbf{M}_i^{-1}\|\|\widehat{\mathbf{u}}_i\|+\|\widehat{\mathbf{u}}_j\|\right) \tag{2.17}$$

$$=o_p(1) \tag{2.18}$$

where the term on the right-hand-side is $o_p(1)$; (2.14) is $\mathbf{S}_{XX}^{-1}+o_p(1)$ as $N\to\infty$ and T fixed; (2.15) is $O_p(1)$ because of $N^{-1}\|\widetilde{\mathbf{X}}_l\|=O_p(N^{-1})$ for $l\in\{i,j\}$, $\|\mathbf{M}_i^{-1}\|=O(1)$ by (A.37), and $\|\mathbf{H}_{ij}\|=N^{-1/2}\|\widetilde{\mathbf{X}}_i\|\|(N^{-1}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\|N^{-1/2}\|\widetilde{\mathbf{X}}_j\|=o_p(1)$; component (2.16) is $O_p(1)$ because $\left(\mathbf{M}_j-\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\mathbf{H}_{ij}\right)^{-1}=\mathbf{I}_T+o_p(1)$ since $\mathbf{M}_j=\mathbf{I}_T-N^{-1/2}\widetilde{\mathbf{X}}_j(N^{-1}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}N^{-1/2}\widetilde{\mathbf{X}}_j'=\mathbf{I}_T+o_p(1)$; and, (2.17) is $O_p(1)$ because $\|\widehat{\mathbf{u}}_l\|=\|\widetilde{\mathbf{u}}_l\|+\|\widetilde{\mathbf{X}}_l\|\|(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\|=O_p(1)$ by ASM.4. Therefore, we can conclude that

$$\widehat{\boldsymbol{\beta}}_{(i,j)}=\boldsymbol{\beta}+o_p(1). \tag{2.19}$$

We then show that the estimators $\widehat{\boldsymbol{\beta}}_{(i,j)}$ and $\widehat{\boldsymbol{\beta}}$ have the same asymptotic distribution. Using *Reverse*

Triangle Inequality,

$$\begin{aligned}
&\left\|\sqrt{N}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})-\sqrt{N}(\widehat{\boldsymbol{\beta}}_{(i,j)}-\boldsymbol{\beta})\right\| \\
&\qquad \leq \left\|\left(\frac{1}{N}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\right\|\left\|\frac{1}{\sqrt{N}}\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i\right\| && (2.20)\\
&\qquad + \left\|\left(\frac{1}{N}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\right\|\left\|\frac{1}{\sqrt{N}}\left(\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\mathbf{H}_{ij}+\widetilde{\mathbf{X}}_j'\right)\right\| && (2.21)\\
&\qquad \left\|\left(\mathbf{M}_j-\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\mathbf{H}_{ij}\right)^{-1}\right\| && (2.22)\\
&\qquad \left\|\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i+\widehat{\mathbf{u}}_j\right\| && (2.23)\\
&\qquad = o_p(1)
\end{aligned}$$

where (2.20) is $o_p(1)$; the overall quantity in (2.21) is $o_p(1)$ because the first component is $\mathbf{S}_{XX}^{-1} + o_p(1)$ and multiplies a quantity that is $O_p(1)$ as in (2.14), noting that $N^{-1/2}\|\widetilde{\mathbf{X}}_i\| = O_p(N^{1/r-1/2})$ and, hence, $O_p(1)$ with $r \geq 2$; (2.22) is $O_p(1)$ as in (2.5); (2.23) is $O_p(1)$. As a results,

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}_{(i,j)}-\boldsymbol{\beta}) = \sqrt{N}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) + o_p(1). \tag{2.24}$$

We can now derive the joint distribution of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}_{(i,j)}$. Using *Shwarz Inequality* and *Triangle Inequality*, we get that

$$\begin{aligned}
\left\|\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\beta}}_{(i,j)}\right\| &\leq \left\|\left(\frac{1}{N}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\right\|\left\|\frac{1}{N}\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i\right\| && (2.25)\\
&+ \left\|\left(\frac{1}{N}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\right\|\left\|\frac{1}{N}\left(\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\mathbf{H}_{ij}+\widetilde{\mathbf{X}}_j'\right)\right\| && (2.26)\\
&\left\|\left(\mathbf{M}_j-\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\mathbf{H}_{ij}\right)^{-1}\right\| && (2.27)\\
&\left\|\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i+\widehat{\mathbf{u}}_j\right\| && (2.28)
\end{aligned}$$

where the left-hand-side is $o_p(1)$ following the reasoning used in justifying the order of convergence of Formulae (2.13)-(2.17). Thus, the difference

$$\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i,j)} \xrightarrow{p} \mathbf{0} \quad \text{as } N \to \infty \text{ with fixed } T. \tag{2.29}$$

The exact finite sample variance of $\widehat{\boldsymbol{\beta}}_{(i,j)}$ under the true model assumptions ASM.1–ASM.4 is

$$\begin{aligned}
\mathrm{Var}\big(\widehat{\boldsymbol{\beta}}_{(i,j)}|\widetilde{\mathbf{X}}_i,\widetilde{\mathbf{X}}_j\big) &= \mathbb{E}\Big\{\big(\widehat{\boldsymbol{\beta}}_{(i,j)}-\boldsymbol{\beta}\big)\big(\widehat{\boldsymbol{\beta}}_{(i,j)}-\boldsymbol{\beta}\big)'|\widetilde{\mathbf{X}}_i,\widetilde{\mathbf{X}}_j\Big\} \\
&= \mathbb{E}\Big\{\big(\widehat{\boldsymbol{\beta}}_{(i)}-\boldsymbol{\beta}\big)\big(\widehat{\boldsymbol{\beta}}_{(i)}-\boldsymbol{\beta}\big)'\Big|\widetilde{\mathbf{X}}_i\Big\} && (2.30)\\
&\quad -\mathbb{E}\Big\{\big(\widehat{\boldsymbol{\beta}}_{(i)}-\boldsymbol{\beta}\big)\mathrm{D}_2(.)'\mathrm{D}_1(.)'\Big|\widetilde{\mathbf{X}}_i,\widetilde{\mathbf{X}}_j\Big\} && (2.31)\\
&\quad -\mathbb{E}\Big\{\mathrm{D}_1(.)\mathrm{D}_2(.)\big(\widehat{\boldsymbol{\beta}}_{(i)}-\boldsymbol{\beta}\big)'\Big|\widetilde{\mathbf{X}}_i,\widetilde{\mathbf{X}}_j\Big\} && (2.32)\\
&\quad +\mathbb{E}\Big\{\mathrm{D}_1(.)\mathrm{D}_2(.)\mathrm{D}_2(.)'\mathrm{D}_1(.)'\Big|\widetilde{\mathbf{X}}_i,\widetilde{\mathbf{X}}_j\Big\} && (2.33)
\end{aligned}$$

where $\widehat{\boldsymbol{\beta}}_{(i,j)} = \widehat{\boldsymbol{\beta}}_{(i)} - \mathrm{D}_1(.)\mathrm{D}_2(.)$ with $\mathrm{D}_1\big(\widetilde{\mathbf{X}}_l\big) = \big(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\big)^{-1}\big(\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\mathbf{H}_{ij}+\widetilde{\mathbf{X}}_j'\big)\big(\mathbf{M}_j-\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\mathbf{H}_{ij}\big)^{-1}$, $\mathrm{D}_2\big(\widetilde{\mathbf{X}}_l,\widetilde{\mathbf{u}}_l,\boldsymbol{\beta}\big) = \big(\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i+\widehat{\mathbf{u}}_j\big)$, and $\widehat{\mathbf{u}}_i = \widetilde{\mathbf{u}}_i - \widetilde{\mathbf{X}}_i\big(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\big)$. Under the model assumptions, $N^{-1}\widetilde{\mathbf{X}}_l' \xrightarrow{p} \mathbf{0}$ wich drives $\mathrm{D}_1(.) \xrightarrow{p} \mathbf{0}$ and, therefore, (2.31)–(2.33) are asymptotically zero but $N\mathbb{E}\Big\{\big(\widehat{\boldsymbol{\beta}}_{(i)}-\boldsymbol{\beta}\big)\big(\widehat{\boldsymbol{\beta}}_{(i)}-\boldsymbol{\beta}\big)'\Big|\widetilde{\mathbf{X}}_i\Big\}$ converges in probability to $\mathbf{V}_{\widehat{\boldsymbol{\beta}}}$ due to the only component that does not vanishes towards zero from result (2.9). As a result, the L2O estimator $\widehat{\boldsymbol{\beta}}_{(i,j)}$ has the same limiting distribution as the *within-group* estimator, $\widehat{\boldsymbol{\beta}}$. The distance between $\widehat{\boldsymbol{\beta}}_{(i,j)}$ and $\widehat{\boldsymbol{\beta}}$ hence vanishes as panel units grow and the joint influence exerted by units (i,j) is null.

We derive the joint distribution of $\widehat{\boldsymbol{\beta}}_{(i,j)}$ and $\widehat{\boldsymbol{\beta}}_{(j)}$. Using *Shwarz Inequality* and *Triangle*

Inequality, we obtain

$$\begin{aligned}
&\left\|\sqrt{N}(\widehat{\boldsymbol{\beta}}_{(i,j)}-\boldsymbol{\beta})-\sqrt{N}(\widehat{\boldsymbol{\beta}}_{(j)}-\boldsymbol{\beta})\right\| \\
&\leq\left\|\left(\frac{1}{N}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\right\|\left\{\left\|\frac{1}{\sqrt{N}}\widetilde{\mathbf{X}}_j'\mathbf{M}_j^{-1}\widehat{\mathbf{u}}_j\right\|+\left\|\frac{1}{\sqrt{N}}\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i\right\|\right. \qquad (2.34)\\
&\left.+\left\|\frac{1}{\sqrt{N}}(\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\mathbf{H}_{ij}+\widetilde{\mathbf{X}}_j')\right\|\left\|(\mathbf{M}_j-\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\mathbf{H}_{ij})^{-1}\right\|\left\|(\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i+\widehat{\mathbf{u}}_j)\right\|\right\} \qquad (2.35)
\end{aligned}$$

is $o_p(1)$ by results (2.14)-(2.17). Therefore, $\widehat{\boldsymbol{\beta}}_{(i,j)}$ and $\widehat{\boldsymbol{\beta}}_{(j)}$ have identical asymptotic distribution

$$\sqrt{N}\big(\widehat{\boldsymbol{\beta}}_{(i,j)}-\boldsymbol{\beta}\big)=\sqrt{N}\big(\widehat{\boldsymbol{\beta}}_{(j)}-\boldsymbol{\beta}\big)+o_p(1) \qquad (2.36)$$

which is asymptotically equivalent to $\widehat{\boldsymbol{\beta}}$ by result (2.8) relatively to unit j. Full proof is provided in Appendix A.7.

2.4 Diagnostic Measures for Panel Data

The influence of an observation is evaluated on the basis of the magnitude of the change in the parameter estimates or in the model fit after its deletion. Such a change can be explained by either the high residual or high leverage of an observation, or both depending of the type of anomaly. Diagnostic measures that examine the degree of anomaly of the deleted case are constructed on these two dimensions.

In this section, we formalise statistical quantities for the unit-wise evaluation of the type of anomaly and degree of their influence. Our influence method is built on a unit-wise evaluation of the influence exerted by each unit in the sample and not case by case. We first introduce the notion of average individual leverage and average normalised residual may help the researcher to understand the complexity of the data (i.e., the type of anomaly) before and after the unit's deletion.

To handle multiple atypical cases in a panel framework, we follow Lawrance's (1995) concepts of joint and conditional influence. We hence propose two versions of his diagnostic measures for linear panel data models with fixed effects. These measures quantify the extent of the reciprocal influence that units exert. We provide the proofs in the context of linear panel data models.

2.4.1 Individual Average Leverage and Residuals

The influence exerted by a unit can be evaluated in terms of its leverage and normalised residual squared. In a panel data setting, the overall influence exerted by a unit on the covariate and outcome dimensions needs to take into account of the overall history of the unit. For this purpose, we introduce the notation for the individual average leverage and the average normalised residual squared.

For simplicity, assume $T_i = T$ for all i. The individual average leverage measures the average leverage of unit i over its full history. In formulae,

$$\bar{\bar{h}}_i = T^{-1} \sum_{t=1}^{T} h_{itt} \text{ for all } i = 1, \ldots, N \tag{2.37}$$

with individual leverage $h_{itt} = \widetilde{\mathbf{x}}'_{it}(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{x}}_{it}$. The average normalised residual squared is a measure of the average outlyingness of unit i over time. In formulae,

$$\widehat{u}_i^* = \frac{1}{T} \sum_{t=1}^{T} \widehat{u}_{itt}^* \text{ for all } i = 1, \ldots, N \tag{2.38}$$

with individual normalised residual squared $\widehat{u}_{itt}^* = (\widehat{u}_{itt}/\sqrt{\sum_i \widehat{u}_i^2})^2$. Plotting $\bar{\bar{h}}_i$ over $\widehat{u}_i^*$ informs on the presence and type of anomalous unit in the sample based on their positions on the plane, as described later in Section 2.5.1

2.4.2 Joint Influence

Joint influence is measured by deleting the full history of pairs of units i and j from the sample. In formulae, it is defined as follows

$$\mathrm{C}_{ij}(\widehat{\boldsymbol{\beta}}) = \left(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i,j)}\right)' \left(\sum_{i=1}^{N} \widetilde{\mathbf{X}}_i' \widetilde{\mathbf{X}}_i\right) \left(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i,j)}\right) (s^2 K)^{-1} \approx \mathrm{F}(\nu_1, \nu_2) \tag{2.39}$$

where $\widehat{\boldsymbol{\beta}}$ if the *within-group* estimator of the true population parameter $\boldsymbol{\beta}$; $\widehat{\boldsymbol{\beta}}_{(i,j)}$ is the L2O estimator computed without units i and j in the sample; N is the total number of cross-sectional units in the sample; $\nu_1 = K$ is the total number of regressors including the constant term, and ν_2 is the number of degrees of freedom at the denominator[4]; $s^2 = \nu_2^{-1} \sum_{i=1}^{N} \widehat{\mathbf{u}}_i' \widehat{\mathbf{u}}_i$ is the variance of the fitted model (i.e., residual MSE) and is consistent for σ^2. The diagnostic measure follows a F distribution, $\mathrm{F}(\nu_1, \nu_2)$, with ν_1 degrees of freedom at the numerator and ν_2 degrees of freedom at the denominator.

The estimators $\widehat{\boldsymbol{\beta}}_{(i,j)}$ and $\widehat{\boldsymbol{\beta}}$ share the same asymptotic distribution (proof is provided in Section A.7.1). That is,

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{(i,j)} - \boldsymbol{\beta}\right) = \sqrt{N}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) + o_p(1) \tag{2.40}$$

When $j = i$, Formula (2.39) collapses to a readapted version of Banerjee and Frees's (1997) distance[5]. That is,

$$\mathrm{C}_{ii}(\widehat{\boldsymbol{\beta}}) = \left(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i)}\right)' \left(\sum_{i=1}^{N} \widetilde{\mathbf{X}}_i' \widetilde{\mathbf{X}}_i\right) \left(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i)}\right) (s^2 K)^{-1} \approx \mathrm{F}(\nu_1, \nu_2) \tag{2.41}$$

where $\widehat{\boldsymbol{\beta}}_{(i)}$ is the the Leave-One-Out (L1O) estimator, $\widehat{\boldsymbol{\beta}}_{(i)} = \widehat{\boldsymbol{\beta}} - \left(\widetilde{\mathbf{X}}' \widetilde{\mathbf{X}}\right)^{-1} \widetilde{\mathbf{X}}_i' \mathbf{M}_i^{-1} \widehat{\mathbf{u}}_i$, computed

[4] The degrees of freedom at the denominator are $(NT - N - K)$ when units are non-clustered and $(N - 1)$ when units are clustered.

[5] Banerjee and Frees's (1997) distance is originally developed for RE models. It can be considered a panel version of Cook's (1979) distance.

without observations i (derivation in Appendix A.1). $\mathrm{C}_{ii}(\widehat{\boldsymbol{\beta}})$ measures the influence of unit i on the FE estimate of β. Formula (2.41) can be interpreted as the "conventional" Cook's distance for panel data. As shown, it is a special case of the more general measure for the deletion of pairs of units, $\mathrm{C}_{ij}(\widehat{\boldsymbol{\beta}})$. Using the formula of the L1O estimator $\widehat{\boldsymbol{\beta}}_{(i)}$, Equation (2.41) can be rewritten as

$$\begin{aligned}\mathrm{C}_{ii}(\widehat{\boldsymbol{\beta}}) &= \frac{1}{s^2K}\widehat{\mathbf{u}}_i'\mathbf{M}_i^{-1}\mathbf{H}_i\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i \\ &= \frac{1}{K}\mathbf{r}_i'\mathbf{M}_i^{-1/2}\mathbf{H}_i\mathbf{M}_i^{-1/2}\mathbf{r}_i \end{aligned} \tag{2.42}$$

where $\mathbf{r}_i = \mathbf{M}_i^{-1/2}\widehat{\mathbf{u}}_i s^{-1}$ are the studentised FE residuals, and $\mathbf{M}_i^{-1/2}$ is the inverse square-root matrix obtained through spectral decomposition[6] of $\mathbf{M}_i$. We obtain a similar expression to Huh's (1993) formula (2). These two diagnostic measures are not test statistics but the knowledge of their empirical distribution can be used to extrapolate cut-off values to assess the joint and individual influence of units.

The effect of subject j on the joint influence after the effect of subject i is measured by the ratio between $\mathrm{C}_{ij}(\widehat{\boldsymbol{\beta}})$ and $\mathrm{C}_{ii}(\widehat{\boldsymbol{\beta}})$. We refer to this quantity as $\mathrm{K}_{j|i}$. Large values of $\mathrm{K}_{j|i}$ (above 1) indicate that unit j has an *enhancing effect* on $\widehat{\boldsymbol{\beta}}$ with respect to unit i while small values, $\mathrm{K}_{j|i} \in (0,1)$, a *reducing effect* (Lawrance, 1995). This measure is not adequate to compare individual influences arising *before* and *after* the deletion of another subject (Lawrance, 1995). For this, the notion of conditional influence is needed.

2.4.3 Conditional Influence

Because the notion of joint influence does not provide a comparison of individual influences arising both *before* and *after* the deletion of another subject, the use of a measure that detects the influ-

[6] The spectral decomposition of $\mathbf{M}_i$ is calculated as follows. Let X and L be respectively the orthogonal matrix with eigenvectors and the vector with eigenvalues of $\mathbf{M}_i$. The spectral decomposition of $\mathbf{M}_i$ is $\mathbf{M}_i^{-1/2} = X\mathrm{diag}(L^{-1/2})X'$.

ence of unit i after the deletion of unit j becomes more appealing (Lawrance, 1995). Conditional influence captures the influence of unit i *after* the removal of unit j from the sample. Exploiting the concept of conditional influence, the proposed measure for panel data, based on Lawrance's (1995) Fomula (4.1), is as follows

$$\mathrm{C}_{i(j)}(\widehat{\boldsymbol{\beta}}) = \big(\widehat{\boldsymbol{\beta}}_{(i,j)} - \widehat{\boldsymbol{\beta}}_{(j)}\big)' \left(\sum_{\substack{i=1\\ i\neq j}}^{N} \widetilde{\mathbf{X}}'_{i(j)}\widetilde{\mathbf{X}}_{i(j)}\right) \big(\widehat{\boldsymbol{\beta}}_{(i,j)} - \widehat{\boldsymbol{\beta}}_{(j)}\big)(s^2K)^{-1} \approx \mathrm{F}(\nu_1,\nu_2) \tag{2.43}$$

where $\widehat{\boldsymbol{\beta}}_{(j)}$ is the L1O estimator computed without observations j; and $\widetilde{\mathbf{X}}_{i(j)}$ is a $(N-1)\times K$ matrix of regressors without the jth subject after the *within-group* transformation of the data. The value of $\mathrm{C}_{i(j)}(\widehat{\boldsymbol{\beta}})$ is zero for $j=i$. The diagnostic measure $\mathrm{C}_{i(j)}(\widehat{\boldsymbol{\beta}})$ is not a test statistics but the knowledge of its empirical distribution can be used to extrapolate cut-off values to assess the conditional influence of units.

A measure for comparing the effect on the distance of unit i *before* and *after* the deletion of unit j is the ratio between the conditional, $\mathrm{C}_{i(j)}(\widehat{\boldsymbol{\beta}})$, and individual, $\mathrm{C}_{ii}(\widehat{\boldsymbol{\beta}})$, influence. We denote this ratio as $\mathrm{M}_{i(j)}$ and, according to Lawrance (1995), captures the *masking* or *boosting effect* of unit i by unit j. For values of $\mathrm{M}_{i(j)} > 1$, the individual influence of unit i is masked by unit j while boosted when $\mathrm{M}_{i(j)} \in (0,1)$.

2.5 Visual Inspection of Anomalous Points

We simulate a synthetic sample of data points using Bramati and Croux's (2007) simulation design. We generate a regression model as follows

$$y_{it} = \beta_0 + \beta_1 x_{it} + \alpha_i + \varepsilon_{it}, \text{ for all } i = 1,\dots,N \text{ and } t = 1,\dots,T \tag{2.44}$$

$$x_{it} \sim \mathcal{N}(0,1);\ \alpha_i \sim \mathcal{U}(0,20);\ \varepsilon_{it} \sim \mathcal{N}\big(0,1\big), \tag{2.45}$$

where $\beta_0 = 1$ and $\beta_1 = 0$. Equation (2.44) is estimated using the *within-group* approach. We set $N = 100$ and $T = 20$.

We contaminate 10% of the total number of units in the sample N with three different types of anomalous observations: VO, GL and BL. VOs are obtained by adding values drawn from $\mathscr{N}(50,1)$ to the original dependent variable; GL points are generated by replacing the original value of the regressor with values drawn from $\mathscr{N}(10,1)$; and BL points are created by adding values from the distribution $\mathscr{N}(50,1)$ to the original dependent variable and, then, replacing the original value of the independent variable with values from $\mathscr{N}(10,1)$. We distinguish between *block-concentrated* and *cell-isolated* contamination. In *block-concentrated* anomalous observations, half of the time series is contaminated whereas random elements are contaminated in *cell-isolated* atypical cases.

Because we are interested in the simultaneous influence exerted by multiple atypical observations in the sample, we provide examples were all possible combinations of VO, GL and BL points are displayed. For instance, samples that present the following combinations of points: {BL, GL, VO}, {BL, GL}, {BL, VO}, {GL, VO}, {BL, BL}, {GL, GL}, {VO, VO}.

2.5.1 Leverage-Residual Plots

We plot the individual average leverage, $\overline{\overline{h}}_i$, over the average normalised residual squared, $\widehat{u}_i^*$. The 2-way representation helps the detection of each of the three type of anomalous units that occupy a specific space in the plane, as explained below.

Figure C.7 compares a leverage-residual plot produced from a case-wise evaluation (left panel – the conventional "cross-sectional approach") and from a unit-wise evaluation (right panel – our "panel approach") of the observations in the sample. The first three units in Panel A (at the top) are contaminated with cell-isolated anomalous points whereas in Panel B with block-concentrated anomalous units following Bramati and Croux (2007). Unit 1 is designed to be a GL point, unit 2 a VO, and unit 3 a BL point.

The horizontal and vertical red lines are the maximum values of lack of high leverage and residual, respectively. Points above the horizontal line display high leverage whereas points to the right of the vertical line have large residuals. VOs are located to the right of the vertical line and below the horizontal line. GL units (points) are located to the left of the vertical line and above the horizontal line. BL units (points) are positioned to the right of the vertical line and above the horizontal line. The main cloud of points to the left of the vertical line and below the horizontal line constitutes non-influential units (points).

From a visual inspection, we observe that the unit-wise plots already display the average influence evaluation for unit i (i.e., the individual average influence is plotted against the individual average normalised residuals squared) whereas the assessment with the case-wise evaluation is done on the basis of the cloud of points for unit i. Overall, the three unusual units are clearly visible and identifiable in all types of plots. In the left plots, the visualisation at the intersection of the two lines is quite confused and it is difficult to evaluate the presence of another potential influential case. In our proposed representation (to the right), the main cloud of units is confined to the bottom-left corner of the plane. A potentially problematic unit – that does not lay with the main bulk of data points, because its influence may be masked by one of the most influential units – can be detected in the right plot in Panel A.

2.5.2 Network Graphs

In this section, we use network analysis tools to visualise and inspect the overall an bilateral influence of different units in a way that is more efficient than 2-way graphs (for a comparison, see Figures C.8 and C.9). The resulting matrix from computing the quantities $\mathrm{C}_{ij}(\widehat{\boldsymbol{\beta}}), \mathrm{K}_{j|i}, \mathrm{C}_{i(j)}(\widehat{\boldsymbol{\beta}})$, and $\mathrm{M}_{i(j)}$ resembles a *directed* and *weighted* adjacency matrix from graph theory. In a directed graph, the effect of unit i on unit j differs from j to i; in a weighted graph, the intensity of the effect of each unit is different.

Formally, let G be a graph formed by an ordered pair of $(V, \mathbf{g})$, where $V = \{1, \ldots, N\}$ is a set of vertices (or nodes) representing the data points involved in a network of relationships, and $\mathbf{g}$ a real-valued $N \times N$ adjacency matrix with the g_{ij} entry representing the influence of j and i. The total number of nodes is fixed and does not exceed the total number of units in the network. The links show the connection between units, the arrows the direction of the connections, and their width the intensity level of their connections.

Figures C.9-C.15 show the proposed representation of joint and conditional influence. Each graph reports two panels: Panel A for a simulated sample with cell-isolated units and Panel B with block-concentrated units. Each panel displays the *enhancing effect* to the left and the *masking effect* to the right. Graphs on the *enhancing effect* display the interaction of unit j with high joint influence ($\mathrm{C}_{ij}(\widehat{\boldsymbol{\beta}}) \geq 4/N$)[7] with unit i, whose influence is enhanced by the presence of the former (as $\mathrm{K}_{j|i} \geq 1$). Graphs on the *masking effect* show unit j masking the effect of unit i (when $\mathrm{M}_{i(j)} \geq 1$). The size of the nodes is proportional to the overall joint and conditional influence of the unit in the enhancing and masking effect graphs, respectively. The colour (from light to dark blue) of the nodes reflects the degree of the influence of the unit. The darker the node, the more influential the unit is. The width of the links captures the strength of the connection in enhancing or masking the effect. The thicker the link, the stronger the connection. In addition, the colours of the links match with the color of the nodes.

The common patterns from comparing Figures C.9-C.15 can be summarised as follows. Joint and conditional diagnostic measures are quite helpful in detecting leveraged data units and,

[7] Bollen and Jackman (1985) set $4/N$ as cut-off for the Cook's distance to isolate the influential cases from non-influential observations.

in particular, BL units in showing their link with other units. Overall, BL and GL units have the largest joint and conditional influence and contribute to enhancing and masking the effects of even fairly influential units. These units (in all their possible combinations) are hence the central nodes of the network, connecting different clusters of units with large joint influence. BL units display the highest total joint influence and, therefore, enhance the effect of other (even fairly influential) units. When another anomalous unit of any kind in the sample, they mask the effect of other units and its total conditional influence is largest compared to the rest of the units. GL units interact with the rest of the units in the sample as BL points but with a milder effect, especially in terms of the masking effect. VO do not mask the effect of any other unit and their total influence is not so large, by construction. Their bilateral joint influence is enhanced by other anomalous units.

The main difference between cell-isolated and block-concentrated contaminated data points is that the total influence of the anomalous units is larger in the latter case than former (with the exception for VO).

2.6 Empirical Examples

In this section, we conduct our influence analysis to four available country-level data sets – specifically, Acemoglu et al. (2008), Schularick and Taylor (2012), Égert (2016), and Berka et al. (2018). In each replicated paper, we apply our method to detect potential anomalous units with leverage-residual plots, and network graphs displaying the total joint and conditional influence and the direction of the effect. Then, we use PHC-like standard errors – PHC0, PHC3, PHCjk, and PHC6 as defined and tested in Chapter 1 – to document any expected change in the statistical inference with GL points and heteroskedasticity, and no change with VO and BL points.

Whenever we detect GL units in the replicated studies, statistical inference with conventional cluster-robust standard errors is, as expected, over-inflated with respect to jacknife-based standard errors. However, the presence of BL points or VOs does not affect the statistical inference but the estimated regression coefficients such that cluster-robust standard errors and the hybrid estimator proposed in Chapter 1 coincide.

Finally, the choice of empirical examples is restricted to those studies with a small cross-sectional units given the time span, because it is more likely to have the econometric setting described in Chapter 1. This type of longitudinal data sets are constructed from time series of large aggregates (e.g., countries, regions, etc.) that are common in macroeconomic studies, where the units of observation are generally small.

2.6.1 Example 1: Acemoglu et al. (2008)

Acemoglu et al. (2008) provide evidence for a statistical association between income per capita and various measures of democracy while controlling for other factors finding no cross-country correlation between income and democracy. The econometric model they estimate using the FE approach is

$$d_{it} = \alpha d_{it-1} + \gamma y_{it-1} + \mathbf{x}'_{it-1}\beta + \mu_t + \delta_i + u_{it}, \tag{2.46}$$

where d_{it} is the democracy score of country i in period t, its lagged value is included as a regressor to capture persistence in democracy and also potentially mean-reverting dynamics; y_{it-1} is the lagged value of log income per capita; $\mathbf{x}_{it-1}$ are other lagged controls; μ_t and δ_i are respectively the time and country fixed effects; and u_{it} is the error term.

The replicated analysis uses "Freedom House" data that cover a five and ten-year period spanning over 1960–2000. Data are available for a maximum of 150 countries. We replicate columns (2) and (7) of Table 2. Each specification uses a different time length, from five to ten years. This leads to the loss of the interval of periods corresponding to $t-1$ during the estimation procedure, due to identification and multicollinearity.

We first proceed with the influence analysis to detect the presence of unusual units and their influential behaviour. For each of the replicated specification, we plot the average individual leverage over normalised residual squared in Figure C.16, and then show the total joint and conditional influence and their effects through a network representation in Figure C.17. The leverage-residual plot outlines the presence of several vertical outliers in Specification (2), and the presence of good leverage points and vertical outliers in Specification (7). There are no bad laverage units. The visual inspection suggests that the estimation of the intercept may be compromised in the first model because of VOs while the statistical inference in the second due to the presence of GL points. Looking at the interaction of the most influential units, the network graphs on the left show a cloud of units in the center of the network with high total joint influence (e.g., units 5, 56, 110, and 125 in dark blue and with wider size of the nodes) that enhance the effect of many fairly jointly influential units (in light blue and with smaller size of the node). The graphs on the right clearly show that the same eight units (i.e., 54, 105, 125, 133, 135, 139, 140, and 150) mask the effect of some units, which are fairly conditionally influential. The heteroskedasticity test shows evidence of heteroskedastic disturbances suggesting that HC standard errors are required.

We replicate theit regression results using different formulae of the variance-covariance matrix to calculate standard errors[8] in Table B.6. As expected from the influence analysis, the significance level of the *Democracy* coefficient do not change in Specification (2) when more conservative standard errors are used because VOs do not undermine the statistical inference. It is worth to outline that conventional standard errors are the lowest in magnitude whereas the proposed PHC6 is slightly smaller than the jackknife-type standard errors. In Specification (7), no coefficient is significant with all types of standard errors. However, we observe much smaller conventional standard errors than the other three – which are close in magnitude – suggesting that those three

[8] Original standard errors are not reported in Table B.6 because they are calculated using the asymptotic uncorrected formula of the sampling variance which, by construction, leads to invalid standard errors in the absence of homoskedasticity.

good leverage points may affect the estimates of the variance.

2.6.2 Example 2: Schularick et al. (2012)

The authors are interested in predicting the event of a financial crisis from a country's recent history of credit growth. Annual data for 14 countries is used to estimate the linear probability model with fixed effects is as follows

$$p_{it} = b_{0i} + \sum_{s=1}^{5} b_{1s} \Delta \log(loans/P)_{it-s} + \mathbf{x}_{it}' \mathbf{b}_2 + e_{it}, \tag{2.47}$$

where L is the lag operator; p_{it} is an indicator variable equal to 1 if a financial crisis occurred in country i in year t; $CREDIT$ is the total bank loans deflated by the CPI; $\mathbf{x}_{it}$ contains control factors. We replicate Specifications (2) and (3) of their Table 3. The coefficients in Equation (2.47) are estimated using the *fixed effects* approach for panel data models.

The leverage-residual plots in Figure C.18b display no unusual units, although unit 10 is very close to the threshold for vertical outliers. Figure C.19 displays the connections of the most jointly influential units that enhance the influence of other units. That is, units 7, 9, 10 have a high total joint effect and exert a strong enhancing effect toward other depicted units in Specification (2) while units 8, 9, and 12 in Specification (3). There is no masking effect in the models. The Wald test for groupwise heteroskedasticity confirms the presence of heteroskedasticity, as expected, because we estimate a linear probability model with fixed effects – which generates heteroskedastic disturbances, by construction.

The significance of the regression coefficients in Table B.7 does not vary with the use of different formulae of the standard errors because no leverage unit has been detected with the influence analysis. In both specifications, conventional cluster-robust standard errors are always smaller than the other types of standard errors by 0.01 points whereas PHC6 standard errors either coincide or are slightly smaller than PHC3 and PHCjk.

2.6.3 Example 3: Égert (2016)

Égert (2016) investigates the impact of product and labor market regulations, and the quality of institutions on country-level Multi-Factor Productivity (MFP). They find evidence of a negative impact of anticompetitive product market regulations on MFP, a positive effect of greater openness and higher innovation intensity stimulated by better institutions, on MFP. The econometric regression model is as follows

$$Y_{jt} = \beta_0 + \sum_{i=1}^{n} \beta_n \mathbf{X}_{jt} + \sum_{i=1}^{n} \sum_{l=-k_1}^{k_2} \gamma_{il} \Delta \mathbf{X}_{jt-l} + \varepsilon_{jt}, \tag{2.48}$$

where Y_{jt} is the level of MFP; X_{jit} is a vector of MFP drivers (i.e., innovation intensity, trade openness, adjusted for country size and a measure of product market regulation) for country j at time t; k_1 and k_2 represent respectively leads and lags. Equation (2.48) is estimated with the FE approach[9]. The dataset is an unbalanced panel of 34 OECD countries covering about 30 years at annual frequency. We replicate Specifications (1)–(3) of Table R1 in the Online Appendix and calculating four different types of standard errors.

The influence analysis in Figures C.20 and C.21 highlighted the presence of a unit with possibly high leverage, as well as a bad leverage unit. Figure C.20 plots the average individual leverage against the average normalised residual squared. In all specifications, unit 23 is a BL unit whereas units 10, 17, 14, 28, and 30 are classifiable as VOs. Unit 19 is a good leverage point only in Specification (3) laying slightly below the threshold for high leverage points in the other models. Figure C.21 shows the network interaction of the most influential units in the three specifications. Looking at the plots on the left, units that possess a high total joint influence (units

[9] In the paper, Equation (2.48) is estimated using dynamic OLS.

10, 12, 14, 17) mainly contribute to enhance the effect of the other units (thicker links). Unit 23 (BL unit) appears only in the graph of Specification (3) whereas unit 19 (GL unit) is displayed only in Specifications (1) and (3) but is not that influential. No unit masks the effect of other units in the sample. There is strong evidence of groupwise heteroskedasticity in the data after performing the Wald test for groupwise heteroskedastcity.

From Table B.8 we observe that conventional standard errors are much smaller than the other three formulae whereas PHC6 standard errors are either slightly smaller or equal in magnitude to the two most conservative standard errors. In all three specifications, the level of significance of the variable *Openness size adjusted* switches to a less significant level when PHC6, PHC3 and PHCjk formulae are used. In Specification (3), two more variables, *ETCR public ownership* and *Business expenditure on R&D*, lose level of significance with more conservative standard errors.

2.6.4 Example 4: Berka et al. (2018)

Berka et al. (2018) study the relationship between real exchange rate and sectoral productivity in nine eurozone countries. They find strong correlation between productivity and real exchange rates among high-income countries with floating nominal exchange rates. The estimating regression equation is as follows

$$RER_{it} = \beta TFP_{it} + \mathbf{x}_{it}'\gamma + \alpha_i + u_{it}, \tag{2.49}$$

where RER_{it} is the log real exchange rate (expenditure-weighted) expressed as EU15 average relative to country i (an increase is a depreciation) in period t; TFP_{it} log of TFP level of traded relative to nontraded sector in EU12 relative to country i at time t; $\mathbf{x}_{it}$ are other covariates; α_i is the country fixed effects; and u_{it} is the error term.

The coefficients in Equation (2.49) are estimated using OLS after the *within-group* transformation. The panel sample is balanced and consisting of 9 countries observed over the time period 1995–2007. We replicate columns (2a)–(2c) of their Table 4.

The influence analysis from Figure C.20 highlights the absence of good leverage units and one possible outlying unit with almost twice the average normalised residuals squared (i.e., cut-off value for vertical outliers). However, network interactions based on the joint and conditional influences in Figure C.21 identify unit 14 as highly jointly influential. Figure C.22 outlines the presence of VOs, specifically unit 6 in Specification (2a) and units {3, 8, 9} in Specification (2b). Figure C.23 shows the network representation of the influential units in the three specification. Highly jointly influential units (with darker nodes and thicker links) enhance the effect of units whose total joint effect is weaker (lighter nodes and thinner links). In Specification (2a) the influence of unit 9 is masked by units 6 and 7, where 7 has high total conditional influence and 6 low total conditional influence. There is no masking effect in the other models. There is evidence of heteroskedasticity in the first two specifications but not in the third because the Wald test for group-wise heteroskedasticity test in fixed effects models fails to reject the null hypothesis of constant variance.

From Table B.9, regression coefficients in Specifications (2a)–(2b) are insignificant at all conventional significance levels with all HC formulae of the variance-covariance matrix[10]. However, it is worth to highlight that PHC0 is always around 0.01 points smaller than the most conservative estimators whereas PHC6 ranges between 0.001 and 0.06 points smaller. In Specification (2c), all regressors are insignificant with the exception of $RULC_T$. The significance of the variable $RULC_T$ drops from 1% to 10% level when PHC3 and PHCjk standard errors are used. Unlike the first two specifications, PHC6 standard errors coincide with conventional cluster-robust standard errors suggesting that no unit has a maximal individual relative leverage exceeding the

[10]Original period weighted standard errors are omitted from Table B.9 because they do not account for the presence of heteroskedasticity, by construction.

cutoff value of 2 (see Section 1.6.1) and jackknife-type formulae may be excessively penalising the residuals of the outlying units.

2.7 Conclusion

In this chapter, we developed a unit-wise influence analysis for linear panel data models to identify the type of anomalous unit from their individual, joint and conditional influence. Our contribution is twofold. We formalised the average leverage and average normalised residuals in a panel data setting to produce unit-wise leverage-residual plots. Then, we developed two diagnostic measures for panel data models – based on Lawrance's (1995) cross-sectional measures – showing their statistical distributions. As a novelty, the overall an bilateral influence exerted by different units is displayed through a network graphs.

Overall, we observe that the three types of anomalies can be easily identified in a two-way leverage-residual plot. Joint and conditional diagnostic measures turn out to be quite helpful in detecting leveraged units, and showing their connections with other units. Joint and conditional diagnostic measures are helpful in detecting leveraged data units and, in particular, BL units in showing their link with other units. In particular, network graphs show that BL and GL units have the largest joint and conditional influence and contribute to enhancing and masking the effects of even fairly influential units.

We apply our diagnostic method to four empirical studies to visually detect the presence of any type of anomaly that can invalidate the LS estimates. Once atypical units are detected, we documented that any expected change in the statistical inference with GL points and heteroskedasticity, and no change with VO and BL points.

As a concluding remark, the researcher should not proceed with deletion of any anomalous unit from the sample because each anomaly should be handled accordingly. Two possible scenarios can be outlined. First, with GL points robust statistical inference should be used[11]. Second, with VO and BL robust estimator of the median (e.g., M-estimators, S-estimators, etc.) is available[12].

[11] In these circumstances, our replicated results showed that cluster-robust standard errors are downward biased, unlike jackknife-type standard errors. Because predicted residuals attach a greater penalisation to leveraged data (Hinkley, 1977).

[12] Robust estimation is not further investigated in this thesis. We do not focus on robust estimation techniques because: (i) LS estimation techniques are widely used in cross-sectional and panel frameworks; and (ii) we are interested in analysing the consequences of GL points on the statistical inference and not the effect of BL and VO on the estimated coefficients. As a reference for robust estimators, see: Bramati and Croux (2007), Verardi and Croux (2009), Aquaro and Čížek (2013), Aquaro and Čížek (2013, 2014).

Chapter 3

Gender Sectoral Segregation and Employment Contracts in UK

3.1 Introduction

Over the past decade, the United Kingdom (UK) has adopted several reforms in support of equal treatment of workers in the workplace. This process culminated in 2010 with the Equality Act (EA2010, hereafter) that guarantees equal pay for equal tasks regardless of gender, and sets out several measures prohibiting discrimination in a whole range of areas, such as employment, services and provision of goods[1]. Although the policies led to a more balanced participation rates, gender inequality still remains a persistent issue in terms of the disparity of job opportunities in many occupations and sectors in the UK (Mordaunt, 2019; Kaur, 2020). Women's participation in the traditionally "female-dominated" sectors (such as health-care, food and accommodation, and domestic work) appears to be disproportionately high in the UK (British Council, 2016) like other countries (Bettio et al., 2009; Olivetti and Petrongolo, 2014, 2016; Gomis et al., 2020).

In light of the recent shock that has mainly hit female-dominated in-person jobs, the COVID-19 outbreak has caused further disruption to female labour supply – especially for young

[1] The EA2010 and its related extensions (e.g. Regulations 2011 - Specific Duties and Public Authorities) set out several measures prohibiting discrimination in a whole range of areas, such as employment, services and provision of goods. In this respect, a woman must not be discriminated with respect to a man in a similar situation (*direct discrimination*), or when a particular policy or working practice creates a gender-based disadvantage (*indirect discrimination*).

women, working mothers, and immigrant women – due to additional unexpected childcare and housekeeping work (Czymara et al., 2020; Open Society Foundations, 2020; Johnston, 2021). These facts made Gupta (2020) and Goldin (2022) talk about *she-session*.

A vast body of research has studied the reason for gender segregation at the workplace. First, women tend to self-select in low-paid jobs that offer more flexible and family-oriented contracts (Petrongolo, 2004; Bertrand, 2011; Goldin, 2014; Bertrand, 2020; Morchio and Moser, 2021). Second, they possess a comparative advantage in certain environments in terms of human capital and productivity (Petrongolo, 2004). Third, their preferences for the characteristics of jobs differ from men's (Petrongolo, 2004; Bertrand, 2011; Goldin, 2014; Bertrand, 2020; Morchio and Moser, 2021) as women prefer less competitive and less risky environments (Gneezy et al., 2003; Saccardo et al., 2018). Finally, voluntarily or involuntary discrimination (Petrongolo, 2004) and sexual harassment (Folke and Rickne, 2020) may play a critical role in gender segregation in the labour market. From another perspective, jobs segregation affects women and men's choices to work in some sectors and occupations, thereby distorting labour market participation, wages, and employment contracts (Mumford and Smith, 2008). As a result, gendered behavioural traits and social norms turn out to shape the environment they select themselves into, and future wages for men and women (Reuben et al., 2017).

This study focuses on the role of gender sectoral segregation in labour market outcomes in the UK. We study the extent to which gender sectoral segregation shapes the type of employment contract (i.e., part-time, permanent, remote work, number of weekly working hours) and hourly wages for both men and women. To address this question, we use the Labour Force Survey (LFS) quarterly data for the period 2005-2020.

Our work first compares labour market outcomes of workers in female-dominated sectors with those in male-dominated sectors with similar observed socio-demographic and working characteristics using propensity score matching (PSM). Then, we analyse the contribution of observable and unobservable factors in determining the gap in hourly wages by means of the Kitagawa-Blinder-Oaxaca (KBO) decomposition, and predicted and residual wages from Mincerian regressions. These counterfactual techniques allow us to address the selection into observable and unobservable characteristics.

We find that contractual features that are typical of a specific gender (e.g., part-time by women) are more common in sectors dominated by that group. Also, workers employed in female-dominated sectors are on average paid less than those in male-dominated sectors. We observe that the *coefficient effect* rather than differences in human capital mainly explains the differential wage between men and women in both gender dominated sectors but there still exists an unexplained component. Regarding the importance of unobservable factors, we find that women self-select into low-paid jobs, and at the top of the wage distribution women in female-dominated sectors are always paid less then other workers for reasons other than observable skills.

Our contribution to the literature on gender sectoral segregation. The past literature has extensively investigated the measurement and cumulative effects of occupational and job segregation (Blackburn et al., 1993; Watts, 1992, 1995, 1998; Petrongolo, 2004; Cortes and Pan, 2018; Folke and Rickne, 2020; Scarborough et al., 2021)[2] but sectoral segregation has not been fully addressed. Sectoral segregation happens to be not only a social concern but also an economic issue to be accounted for as it contributes to explain labour market differentials in terms of wages between genders (Moir and Smith, 1979; Campos-Soria and Ropero-García, 2016). Most of these studies typically provide descriptive evidence, and only a limited literature considers job segregation of women in atypical contracts (e.g., Petrongolo, 2004) without distinguishing the effect across sectors. We contribute to the existing literature by looking at the consequences of time-varying gender sectoral segregation on labour market outcomes (i.e., employment contracts and hourly

[2]Other works provide evidence of the key role of vertical segregation in motivating the additional hurdles that women face once in employment – such as the access managerial and decision-making positions and jobs in specific sectors which offer better pays, career opportunities and contractual conditions (Bettio et al., 2009; OECD, 2020; Morchio and Moser, 2021).

wages) based on worker's observable skills and socio-demographics characteristics as well as the features of their workplace.

The novelty of this research is triple. First, we build two indicators that measure the degree and gender-type sectoral segregation (i.e., sectoral dominance and sectoral segregation index) to explain job segregation in terms of differences in employment contracts and differences in hourly wages. Second, we use PSM to estimate the average effect of gender sectoral segregation on various labour market outcomes matching for socio-demographic characteristics and workplace features of the worker. Third, we document the importance of both observable and unobservable factors in determining of the gaps in earnings between men and women in male- and female-dominated sectors. While part of the documented wage gap comes from differences in human capital between men and women (Mumford and Smith, 2008), the remainder still remains unexplained and cannot be ascribed to only observed factors (Booth, 2009). After analysing the contribution of the human capital and workplace features, we investigate selection into observable and unobservable factors in determining hourly wages in female- and male-dominated sectors. We use a methodology taken from the literature of the economics of migration[3] to calculate individual potential wages (Gould and Moav, 2016; Borjas et al., 2019) and capture the part of wages that is uncorrelated to observed skills (Parey et al., 2017)[4].

The rest of the chapter is structured as follows. Section 3.2 describes the data and reports some descriptive analysis. Section 3.3 presents the identification strategies for employment contracts and wages. Section 3.4 reports the estimated results. Section 3.5 concludes.

[3]It is necessary to use this approach because survey data usually lack of behavioural variables in the questionnaires (Booth, 2009). This is the case with the LFS.

[4]This literature highlights that immigrants could be positively/negatively selected with regard to both observed characteristics (e.g., higher levels of education) but also to unobserved determinants of labour market success (e.g. motivation, ambition and ability) that can enter into the decision to self-select into migration (Chiswick, 1978, 1986, 1999; Borjas, 1987; Bertoli et al., 2016).

3.2 Data and Descriptive Statistics

Our analysis is based on the Labour Force Survey (LFS) quarterly data released by the Office for National Statistics (ONS). It is the most extensive household study in the UK providing a comprehensive source of data on workers and the labour market. Our final estimation sample includes the working-age population (aged 16-64) over the fiscal years 2005 and 2020, consisting of 1,788,945 women and 1,544,280 men. As data are already available for the year 2020 (and the first quarter of 2021), this timeliness allows us to take into account also recent significant changes caused by the COVID–19 outbreak. We use the time span going from 2005 to 2020 because the UK labour market has faced several structural changes that are likely to have affected the gender segregation across sectors (e.g., EA2010, the economic crisis in 2008, the COVID-19 outbreak, etc.).

The data include variables on a wide range of: (i) Demographic characteristics (gender, age, nationality, ethnicity, religion); (ii) socio-economic factors (presence of dependent children, marital status, education, experience, full/part-time job, remote work, public sector, training opportunities, sectors and occupations); (iii) geographical information on residence and working region. We distinguish between UK natives and immigrants (male and female), including both citizens from the European Economic Area (EEA) and non-EEA. Salary information in the LFS is the self-reported gross weekly pay for the reference week[5]. We consider 19 sectors of the economy, according to the Standard Industrial Classification (UK SIC) at one-digit[6].

Tables B.10 and B.11 report the summary statistics of the main variables by gender. There is prevalence of natives in both male and female samples(above 80%), followed by non-

[5]We calculate the real wage based on hourly wages in 2015 prices (see UK ONS) as: real wage = hour pay/(CPI2015/100).

[6]Our analysis uses UK SIC 2007, the current five-digit classification used in identifying business establishments by type of economic activity. For years before 2008, we used the correspondence between the sections of SIC 2003 and SIC 2007.

EEA and EEA. The average age for women (mean=39.85; sd=13.52) is smaller and less sparse than those for men (mean=40.14; sd=14.26). On average women in the sample are slightly more educated than men in terms of years of education (for women: mean=13.21; sd=3.05; for men: mean=13.11; sd= 2.87) but with less experience measured in years (for women: mean=23.74; sd=13.27; for men: mean=24.30; sd=13.86). Slightly more than half of the women in the sample are either married or cohabiting whereas exactly half of the men are in a stable relationship. In addition, 37% of the women have dependent children against only 28% of men. The average number of hours worked by women per week[7] (mean=30.89; sd=13.36) is much smaller than male's figures (mean=40.33; sd=13.56). A higher share of women working part-time (43% against 12% for men) affects the average above. A more detailed investigation on the reason for part-time work among women in Table B.12 shows that 8% of working women could not find a full-time job whereas 57% chose to work part-time, mainly for family and domestic commitments (about 46% of women in the sample).

3.2.1 On the Entry Decision

A woman's working decision is made either fully jointly with her partner or conditional to her partner's labour market choices (Goldin, 2006). In the second scenario, women act as a "precautionary earner" to insure the household against the higher risk of unemployment of her partner in recessions (Ellieroth et al., 2019). A preliminary descriptive analysis focuses on the factors that drive the decision to enter the labour market, by contrasting men and women. We estimate the correlations of being in the labour force on a set of socio-demographic, economic and regional variables by means of a Probit model.

Table B.15 reports the marginal effects of the Probit model by gender. Columns (1)

[7] Whenever applicable, the number of hours includes usual hours of paid overtime to total usual hours worked in main job.

and (3) are the estimates from the full sample (2005-2020) and Columns (2) and (3) from the 2020 sample. With respect to British-natives, European men are more likely to be active unlike European women (whose signs are negative. In contrast non-European men and women are less likely to be active, although the magnitudes are higher for women than men, as expected. Women in a long-term relationship (either married or in a civil relationship) tend to be out of the labour force with a probability of 3.4 percentage points (p.p.) in the full sample and 4.3 p.p. during COVID-19 in stark contrast to men's figures that are positive. The presence of dependent children on average increases the likelihood of entering the labour market by 6.5 p.p. for men and only 2.3 p.p. for women over the full period in analysis. During the COVID-19 the magnitudes for women increase to 4.3 p.p. Overall, figures for women support the empirical evidence of a reduced "child penalty" on mother's labour supply over the past decades (Boushey et al., 2005; Goldin, 2006). Compared to individuals with low education, more educated people are less likely to be in the labour force but the magnitudes are quite small. In addition, receiving benefits of any kind decreases the probability to be in the labour force by around 23-25 p.p. for both men and women.

3.2.2 Sectoral Segregation

Sectoral gender segregation arises when there exists a disproportionate share of men and women within and across sectors with respect to total employment – independently of the nature of the job allocation (Watts, 1998). We define a sector to be *female dominated* if the share of women employed in that sector is higher than the corresponding share of men in that sector; it is *male dominated* otherwise. In formulae, sectoral dominance is defined as follows:

$$\text{Sectoral Dominance} = \begin{cases} \text{Female} & \text{if } \frac{W_{jt}}{W_t} > \frac{M_{jt}}{M_t} \\ \text{Male} & \text{otherwise} \end{cases} \tag{3.1}$$

where j is the sector (SIC 1-digit) and t is time frame; W_{jt} and M_{jt} are respectively the total number of women and men employed in sector j at time t; W_t and M_t are respectively the total number of female and male workers at time t. Female and male-dominated sectors are listed in Table B.13.

The classification criterion provided in (3.1) to distinguish between male (md) and female (fd) dominated sectors is used to construct a new measure of gender concentration across sectors. Based on the proportion of men and women in a specific sector, the Sectoral Segregation Index (SSI) measures the degree of the disproportion in the distributions of men and women within and across sectors[8]. Our index informs on the proportion of women that would have to either leave or enter each sector not to have segregation. In formulae, the index is as follows:

$$SSI_t^s = \frac{1}{2} \sum_{j \in J_s} \left| \frac{W_{jt}}{W_t} - \frac{M_{jt}}{M_t} \right| \quad \text{for all } t \text{ and } s \in \{md, fd\} \tag{3.2}$$

where W_{jt} and M_{jt} are respectively, the total number of women and men employed in sector j at time t; W_t and M_t are respectively, the total number of female and male workers at time t. For female dominated sectors, $j \in J_{fd}$ (left side of Table B.13); for male dominated sectors, $j \in J_{md}$ (right side of Table B.13).

The index ranges between 0 and 1 for each sector; the higher the index, the greater the gender sectoral segregation. For each time period t we calculate two sectoral indices: SSI^{fd} is the value of the index for female dominated sectors and SSI^{md} for male dominated sectors. The value of the index remains unchanged when transferring workers between sectors (SIC1) within each group (male and female) in a given gender dominated sector. Nevertheless, SSI is constructed to change when the transfer is across groups in a specific gender dominated sector[9]. Hence, the

[8] Our index is constructed following the Index of Dissimilarity (ID). The ID is a well-established measure in labour (Watts, 1998) and education literature (Zoloth, 1976; James and Taeuber, 1985) to study group composition and quantify the segregation among two groups (such as demographic minority and non-minority groups). Specifically, it provides information on the proportion of the minority group that would have to be transferred to reach no segregation (Cortese et al., 1976; Zoloth, 1976; Watts, 1998).

[9] A similar interpretation is provided by Zoloth (1976) to describe the racial composition of schools within and across districts.

index does not inform on the direction of the gender disproportion because it is always positive by construction, and it is not possible to identify which group drives the unbalance. However, the classification criterion (3.1) allows us to check whether a sector is unbalanced towards men or women. The information provided by the *SSI* can be used to identify sectors with high or low gender segregation[10]. Table B.14 shows high and low segregated sectors according to our classification.

The densities of SSI^{md} and SSI^{fd} are displayed Figures C.25 and C.26 over the entire period of study (respectively, solid and long-dashed lines) and after EA2010 (respectively, dotted and short-dashed lines). Looking at the support of the index in Figure C.25, the aggregate gender segregation in the UK labour market is relatively small in both male and female-dominated sectors. The maximum index level is 0.174 in female-dominated sectors in both time samples, while for male-dominated sectors, 0.18 in the full-time sample and 0.173 after 2010. From a visual comparison of the two total sample distributions, the bulk of the mass of female-dominated sectors is around its peak at 0.17. In contrast, the density of male-dominated sectors is spread over more extensive support and is bimodal at 0.161 and 0.173. When values of the index before the reform are excluded, there is a shift to the left of the support for male-dominated sectors. This means that sectoral gender segregation has reduced after EA2010. However, the distribution in male-dominated sectors remains bimodal but with a greater density around lower levels of the index (around 0.162) confirming a reduction in segregation, unlike in female-dominated sectors where higher levels of segregation are registered (with a peak at around 0.17).

Figure C.26 distinguishes between male and female-dominated sectors with high and low gender segregation. Among low segregated sectors (left panel), gender segregation appears to be smaller in male-dominated sectors, although the density around the right peak of its distribution

[10] Sectors are ranked from the least to the most segregated by gender dominance as follows: a sector is classified as low gender segregation if it is ranked, on average, below the mean rank; a sector is classified as high gender segregation if it exceeds the mean rank on average.

increases after 2010. The index distribution in female-dominated sectors is skewed to the left with a thick right-tail in the total sample but after 2010, the density around the peak increases. The distribution shifts slightly to the left, meaning that gender segregation in female sectors decreases after the reform. Among high segregated sectors (right panel), the distribution of SSI in male-dominated sectors is skewed to the right in the total sample but evenly distributed after 2010. On the contrary, gender segregation in female-dominated sectors is, on average, smaller in terms of magnitudes, but the index's distribution shifts upwards after EA2010.

The trend highlighted in the graphs presents two plausible scenarios: the UK labour market may have experienced either a higher inflow of women into male-dominated sectors (in this case, the Equality Act 2020 had a positive effect) or a higher transition of men into unemployment. We decompose the overall effect using a shift-share sectoral analysis to shed light on these scenarios.

3.2.3 Shift–Share Decomposition of Employment

The evolution of labour market outcomes (employment, unemployment and inactivity rates) between 2005 and 2020 is shown in Figure C.24 by gender. Female employment rate remained stable at around 55% until 2010. After EA2010, it started a gradual increase until the pandemic outbreak in 2020, where it settled to the level reached in 2018 (58%). Similarly, the inactivity rate displays a plateau at about 42% before the reform, showing a decreasing trend after the entry into force of EA2010. The female unemployment rate is relatively low over the period and quite similar in percentages to the male ones, unlike the employment and inactivity rates, which are respectively lower and higher.

To better understand the determinants of the change in male and female employment shares, we adopt a revised version of Olivetti and Petrongolo's (2016) shift-share decomposition[11].

[11] Unlike the original paper that uses the number of worked hours, we use the employment share.

The growth of female and male employment share is decomposed into two components: a first component captures the change in the total *employment share* of the sector (*between component*); a second component reflects changes in *gender composition* within the sector (*within component*). The shift-share decomposition is defined as follows:

$$\Delta e^f_{st} = \underbrace{\sum_{j=1}^{J_s} \alpha^f_{jt} \Delta e_{jt}}_{\text{Between-sector}} + \underbrace{\sum_{j=1}^{J_s} \alpha_{jt} \Delta e^f_{jt}}_{\text{Within-sector}} \quad \text{for all } s,t \tag{3.3}$$

where $\Delta e^f_{st} = \frac{E^f_{st}}{E_{st}} - \frac{E^f_{st_0}}{E_{st_0}}$ is the difference in the share of female employment between the base time period t_0 and the current time period t; $\Delta e_{jt} = \frac{E_{jt}}{E_t} - \frac{E_{jt_0}}{E_{t_0}}$ is the difference in the share of total employment in sector j between t_0 and t; $\Delta e^f_{jt} = \frac{E^f_{jt}}{E_{jt}} - \frac{E^f_{jt_0}}{E_{jt_0}}$ is the difference in the share of female employment in sector j; $\alpha^f_{jt} = \frac{(e^f_{jt_0}+e^f_{jt})}{2}$ and $\alpha_{jt} = \frac{(e_{jt_0}+e_{jt})}{2}$ are decomposition weights: the average share of female employment in sector j and the average share of sector j, respectively. The reference year is the first available year in the dataset ($t_0 = 2005$); s stands for sectors classified as female/male dominated according to Equation (3.1).

Figures C.27 displays the shift-share decomposition of female and male employment (respectively, at the top and bottom). Both graphs show the difference in employment in the comparison year with respect to the base year (i.e., the fiscal year 2005) for women (at the top) and men (at the bottom). The overall change in employment is shown in solid line and its decomposition into the *between* and *within* components respectively, with dashed and dotted lines. The cross marks the components for female-dominated sectors and the circle the components for male sectors. In this way, we can investigate which term drives the overall change in employment and assess the effect of economic downturns and policies on sectors with similar characteristics.

Female composition (*within* component) in the top graph started to gradually increase in male-dominated sectors after the EA2010, whereas it suddenly increased in female-dominated sectors after the economic crisis in 2008 but then decreased after 2012. As expected, the *between* and *within* components in female-dominated sectors drop in 2020 due to the pandemic outbreak. In contrast, there was a rapid rise in female employment in male-dominated sectors. Total employment shares in female-dominated sectors (*between* component) were almost close to the levels of the base year until 2008; after that there was a rise in female employment in female-dominated sectors that was arrested by the COVID–19 outbreak. These results are in line with previous literature on the theme (Hoynes et al., 2012; Doepke and Tertilt, 2016; Ellieroth et al., 2019; Alon et al., 2020).

The composition of male employment remained unchanged in female-dominated sectors. As expected, the 2007-2009 crisis hit male-dominated sectors more than female-dominated sectors. After the EA2020, male composition in male-dominated sectors decreased and remained stable in female-dominated with respect to 2005. The COVID–19 outbreak did not arrest total male employment with respect to previous years.

Overall, the shift-share decomposition highlights interesting facts. First, the 2007-2009 crisis harshly hit male-dominated sectors while stimulating female employment[12] in female-dominated sectors. On the contrary, the COVID–19 outbreak arrested the overall employment in both male and female-dominated sectors and led to a reduction in female employment in female-dominated sectors, in stark contrast to male-dominated sectors. In addition, the Equality Act 2010 did stimulate female employment from the demand side, as we observe a substantial increase in female composition in male-dominated sectors after 2010. This means that a higher proportion of women were employed within each male-dominated sector at the expense of decreasing male employment (the contrast is visible with the bottom graph of Figure C.27).

[12] Similarly, Ellieroth et al. (2019) finds that married women are more stuck in employment during recessions. Therefore, their labour supply decisions account for the higher risk of job loss experienced by their husband.

3.3 Methods

In this section, we present the methods used to evaluate the contribution gender-based segregation on employment contracts and wages, based on observable and unobservable factors. We first implement a Propensity Score Matching (PSM) method to quantify the average difference in labour outcomes (i.e., permanent jobs, part-time jobs, working hours, remote work, and hourly wages) in male- and female-dominated sectors, matching on observable skills and socio-demographic characteristics. Then, we further inspect the differences in hourly wages taking into account worker's observable and unobservable characteristics. For this purpose, we first conduct the counterfactual 3-fold KBO decomposition to examine the components that mainly drive the wage gap in sectors and, then, run a Mincerian wage regressions to explore associations with human capital variables. As workers may self-select on observable and unobservable characteristics into the labour market, we show the trajectory of predicted and residual wages from Mincerian regressions.

3.3.1 Differences due to Observable Factors

Because workers may self-select into the labour market based on their observable skills and human capital, the PSM method accounts for the selection bias matching on observable factors. Technically, the matching is done on the propensity score – i.e., the conditional treatment given the covariates, $P(D_{it} = 1|\mathbf{X}_{it})$. Controlling for the propensity score eliminates the selection bias while controlling for observed factors (Cameron and Trivedi, 2005, pp. 872-873). We implement matching estimators *à la* Abadie and Imbens (2006) to compare the average differences in labour market outcomes (i.e., types of employment contracts and wages) in male- and female-dominated sectors among workers with similar characteristics[13]. Working in female-dominated sectors is used as the

[13]The choice of covariates for the PSM is based on the relevant literature on gender segregation and the model selection performed by LASSO. The choice among selected covariates from the penalised regressions is reported in Table B.19.

treatment status. The underlying assumption is that workers who choose to work in female- and male-dominated sectors only differ in the endowment of their observed skills and revealed preferences. We conduct a standard sensitivity analysis to check the common support and the balancing property of the covariables before and after matching in the treated and non-treated groups. The included covariates are balanced if the standardised bias after matching is within $\pm5\%$ (Rosenbaum and Rubin, 1985). If the condition is satisfied, the matching method successfully builds a meaningful control group.

Further analysis is conducted for differences in wages to investigate whether differences in human capital or discrimination can explain the pay gap. In doing so, we proceed in two stages. First, we conduct the KBO decomposition to examine how these differences in hourly earnings are related to a component accounted for by differences in human capital characteristics and an unexplained component. The KBO decomposition (Kitagawa, 1955; Blinder, 1973; Oaxaca, 1973) is used in the discrimination literature to study outcome gaps between two groups[14]. The outcome differential (here, hourly wages) between two groups (male and female workers) is divided into a part that is explained by observable group differences in productivity and background characteristics (*endowment effect*) and a residual component that cannot be explained by such observed differences in the outcome variable (*coefficient effect* – i.e., discrimination) (Jann et al., 2008). Specifically, when the *endowment effect* is negative, female workers possess better predictors (i.e., characteristics) than their male counterpart. When the *coefficient effect* is positive, discrimination towards women the explain the wage gap.

Then, we use a Mincerian regression to analyse how women's human capital characteristics – i.e. years of education, experience and training opportunities – are associated with hourly wages between sectors and genders. The Mincerian wage regression estimates the correlations of

[14]The derivation of KBO decomposition can be found in Appendix A.8.

control variables (including socio-demographic, human-capital and work-related variables[15]) on hourly pay in logarithm using OLS.

3.3.2 Differences due to Self-selection on Observable and Unobservable Factors

Wage differentials between men and women in male- and female-dominated sectors may also arise due to worker's self-selection on observable and unobservable factors. These may eventually affect trends in observed wages (Bertrand and Hallock, 2001; Blau and Kahn, 2017). In this section, we show the trajectory of potential and residual wages due to self-selection on observable and unobservable characteristics. We use a methodology usually adopted in the literature on migration for selection[16] (Gould and Moav, 2016; Parey et al., 2017; Borjas et al., 2019).

Because the literature shows that women tend to self-select into the labour market not only on observable skills but also on unobservable characteristics – such as, career ambitions, competitiveness, bargaining power etc. (Gneezy and Rustichini, 2004; Booth, 2009; Bertrand, 2011), it is important to further investigate this aspect. Therefore, we show the distribution of predicted wages and residual wages from Mincerian regressions, by gender-sectoral dominance and across genders. The former captures the potential wage of the two groups based on observable characteristics while the second reflects the part of wages that is uncorrelated with observed skills and human capital.

These two measures are then used to construct the Cumulative Distribution Functions (CDF) by gender and sectoral dominance. We then compare the CDFs of predicted and residual

[15] Socio-demographic variables include: age and its square, nationality, ethnicity, religion, being in a stable relationship, having dependent children and the interaction of the last two. Human-capital variables are education, experience and its square, years in education and its square, training offered by current employer. Work-related variables include the type of occupation. Working region dummies are included. Household income is not provided and cannot be constructed with the EUL version of the data.

[16] With survey data it is difficult to proxy the psychological behaviour (Booth, 2009), and the unexplained component of the KBO decomposition always remains unexplored (Mumford and Smith, 2008).

wages between men and women. We determine whether the distributions of the predicted wages and residuals statistically differ among the two groups by performing the Kolmogorov-Smirnov (K-S) test.

3.4 Results

This section discusses the empirical results of the conducted analysis. With the PSM, we observe that gender sectoral segregation does contribute to differences in employment contracts and hourly earnings. In female-dominated sectors workers are paid less, work more part-time and, hence, less hours, and less remotely compared to workers in male dominated sectors. With the KBO decomposition, we show that the gap is mainly explained by the *coefficient effect* rather than differences in endowments (i.e., human capital). In addition, we find that women are negatively selected at the top of the wage distribution, unlike men, especially in female-dominated sectors, where they are paid less for reasons other than their skills.

3.4.1 Differences due to Observable Factors

Tables B.21-B.25 summarise the main results of the PSM[17], where working in a female-dominated sector is the treatment. The Average Treated Effect on the Treated (ATET) for several types of employment contracts is as follows: 0.011 for having a permanent job; 0.135 for part-time work; -0.125 for log hours worked; and -0.044 for remote work. The results suggest that contractual features that are systematically chosen by a specific gender[18] tend to be predominant in sectors dominated by that dominant group. The ATET for log hourly wage is -0.094, implying that workers in female-dominated sectors are on average paid 9.4 percentage points (p.p). less than

[17]The estimates in Table B.20 from a Probit regression are used to calculate the propensity score for the PSM.

[18]For example, Petrongolo (2004) provides descriptive evidence that European women are more likely to segregate in atypical jobs (e.g., part-time or temporary jobs)

those working in male-dominated sectors based on matched observed characteristics. This result is aligned with Folke and Rickne (2020), who find that female-dominated workplaces are made less attractive (for men) in terms of both monetary and non-monetary standpoint.

We conduct a sensitivity analysis to check whether the balancing property and common support of the covariates are satisfied. Figure C.30 and Table B.26 reports the single components of the covariate imbalance test. Almost all covariates are well balanced – with standardised bias after matching between $\pm 5\%$. Overall, the matching method effectively built a valid control group. The chosen matching algorithm imposes a common support such that treatment observations whose propensity score is greater than the maximum or lower than the minimum propensity score of the controls are dropped. The number of dropped treated units off support are 16. The number of untreated and treated units on the support are reported underneath each table.

Then, we look at the association between human capital variabels and hourly wages by means of the Mincerian earnings regression. Table B.28 reports the estimated coefficients of the Mincerian wage regression[19] for male and female workers, pooling the sectors (columns (1)-(2)) and by gender-dominated sectors (columns (1)-(6)).

Looking at socio-demographic characteristics, age is positively correlated to higher wages, despite the small magnitudes, and the relationship is quadratic. European (EEA) and non-European (non-EEA) workers earn on average less than British-Natives in all samples. However, the reduction in magnitudes is on average higher for EEA than non-EEA, for EEA in male-dominated sectors but for non-EEA in female dominated sectors. The presence of dependent children has a strong negative correlation on women's wages in all samples while the correlation is positive for men. The negative correlation is overall attenuated for married women with dependent children (2.0 p.p. in the full sample) with high and significant magnitude in male-dominated

[19] Usual worked hours per hour and its square are not included in the regression specifications because of possible endogeneity issues due to reverse causality. In addition, because hourly wages are calculated based on usually worked hours per week estimates will be downward biased due to the division bias (Borjas, 1980).

sectors (4.5 p.p.); there is no significant correlation for men.

We now comment on human capital variables. Focusing on educational attainment[20], having intermediate or high education is positively associated with wages unlike low education; magnitudes are slightly higher for women than men for high education in all samples. As expected, more years of education increase wages but with a diminishing effect (the square is negative). Looking at the estimates for years of education, the optimal number of years in education that maximises wages for men is approximately 15.8 years ($= 0.158/(2 \times 0.005)$) while for women it is about 19.5 years ($= 0.117/(2 \times 0.003)$) in the full sample. Consequently, women are required to have higher education, unlike men who need just a degree to earn their optimal wages. The difference is driven by female-dominated sectors ($17.3 = 0.104/(2 \times 0.003)$ for women and $16.3 = 0.130/(2 \times 0.004)$ for men) because the optimal wage is 16.5 years for both men and women in male dominated sectors. Potential working experience has significant diminishing returns (the coefficient of experience is positive and its square negative but very small) and receiving a training increases the hourly wage, especially in male-dominated sectors.

On the workplace characteristics, the association of women working in public sector with hourly wages is, on average, 5.6 p.p. higher than those in private sector and the magnitude is higher than men's. However, the correlation is negative in male-dominated sectors but positive in female-dominated sectors both men and women. This suggests that private sector pays more in male-dominated sectors while the public sector offers better remuneration for female-dominated sectors. As expected, working part-time is negatively correlated with hourly wages; magnitudes are higher for men than women suggesting a higher penalty for men. Working in sectors with low gender sectoral segregation significantly increases the association with wages of male workers on average by 2.7 p.p. when pooling sectors, by 2.4 p.p. in male-dominated sectors and 5.9

[20]In the Mincerian regression we included both the categorical variable for education band (low, intermediate, and high education) and the continuous variable for years of education and its square. The OLS assumption of absence of perfect multicollinearity is not violated because years of education capture the intensity of the returns of education within an education band. The information provided by the two variables is hence complementary.

p.p. in female-dominated sectors. Conversely, low gender sectoral segregation does not have a significant correlation with women's wages in the pooled sectors sample and male-dominated sectors but in female-dominated sectors wages are expected to be 3.4 p.p. higher then sectors with high segregation. Hourly wages in female-dominated sectors are on average lower than in male-dominated sectors (correlations are 16.3 p.p. for men against 15.8 p.p. for women). These figures support the claim that female-dominated sectors pay on average less for both men and women, as per the PSM results. The interaction term between female-dominated sectors and low gender segregation is positive and significant for women only (as already observed in column (6) for the coefficient on low gender segregation).

We then decompose the average difference in log wages between men and women based on the Mincerian regression model to study the labour market compensation in a counterfactual manner. The evolution of the three components of the KBO decomposition and their sum are shown in Figures C.28 (full sample) and C.29 (by gender sectoral segregation) over the full period of study. The dashed line represents the *coefficient effect*, the long-dashed line the *endowment effect* and the dotted line the part of the "unexplained" component of the three-fold decomposition (or *interaction effect*). The corresponding shadowed areas display the 95% confidence intervals. The solid line is the sum of the three effects and reveals their overall contribution. The contribution of main socio-demographic characteristics, human capital attributes and sectoral indicators are displayed in Tables B.16, B.17 and B.18.

In Figure C.28, the difference in wages between men and women is on average of 0.2 logarithmic points over time. Most of the gender pay gap (around three fourths) can be explained by differences in the estimated coefficients between genders. In fact, the *coefficient effect* on average quantifies an increase of 0.16 logarithmic points in women's wages when the male coefficients are applied to female characteristics. In addition, this component displays a downward trend after 2008. The *endowment effect* quantifies an expected average increase in women's wage by around

0.05 points if they had male predictors levels. Therefore, differences in observed characteristics account for one fourth of the gap.

When we repeat the exercise by gender sectoral dominance in Figure C.29, the *coefficient effect* is still positive but steadily much higher in male-dominated sectors (around 0.2 points on average overt time) than in female-dominated ones (about 0.13 points on average and decreasing over time). This suggests that women should be paid more than men to prevent any form of *discrimination effect* between the two groups. The dynamics of the *endowment effect* differ in both male and female-dominated sectors. Specifically, women working in male-dominated sectors should expect an average decrease in their wage by a small amount if they had men's predictors before 2010 and after 2018 but no change in the period within this window (the *endowment effect* is slightly positive but close to zero). Conversely, men and women working in female-dominated sectors are always similar in terms of human capital as the *endowment effect* is on average around zero. The *unexplained component* has a stronger negative contribution in male-dominated sectors after 2010, capturing all the potential effects of differences in unobserved factors other than human capital that contribute to shaping the trajectories of wages in these sectors – such as, self-esteem, ambition, bargaining power, etc. (Booth, 2009). Overall, the *coefficient effect* prevails over the other two, despite being partly offset by the negative *unexplained effect* in male-dominated sectors.

3.4.2 Differences due to Self-selection on Observable and Unobservable Factors

This section presents empirical evidence on the differences in average hourly wage between men and women when considering unobservable characteristics of workers. Figures C.31 and C.32 display the Cumulative Distribution Functions (CDF) of the predicted earnings (on the left) and residual earnings (on the right) when pooling all sectors and distinguishing male- and female-dominated sectors, respectively. The solid lines are for men and the dashed lines for women.

From a visual inspection, men are positively selected in terms of earnings potential once in the labour market because the CDFs for men are to the right of the CDFs of women in the left graphs. Positive selection on unobserved characteristics implies that wages cannot be attributed to acquired skills and/or human capital only. That is, men seems to have abilities or qualities that are valued more than those of women, according to the Mincerian regressions. In addition, the CDFs for men and women do not cross, which means that there is positive selection over the full support of predicted earnings. From a comparison of CDFs in Figure C.32, predicted earnings of women (men) in female-dominated with those of women (men) working in male-dominated sectors are lower. This result supports Folke and Rickne's (2020) claim that female-dominated workplaces are made less attractive for men from a monetary (and non-monetary) viewpoint while male-dominated sectors offer higher remuneration, despite being less attractive to women.

Focusing on graphs on the right, the CDFs of the residual earnings of women lie to the right of the CDFs for men for low values of residual hourly earnings whereas to the left for high values. Because the two CDFs cross at zero, this implies that women are positively selected in terms of wage residuals at the bottom of the distribution, unlike men who are positively selected at the top of the distribution. When looking at the graphs by gender-sectoral dominance, we observe that the CDFs are almost perfectly aligned in male-dominated sectors with the CDF of women to the right of men's at the very top of the wage distribution. The distinction between the two curves is more evident in female-dominated sectors where women self-select into low paid jobs. Overall, it seems that male and female workers in male-dominated sectors are similar in terms of observed and unobserved characteristics such that the difference in the trajectory of earnings is residual.

The Kolmogorov–Smirnov (K-S) test checks whether the two data samples come from the same distribution. The combined K-S test statistic for predicted wage is $t_{\hat{y}} = 0.2067$ and for residual wage $t_{\hat{u}} = 0.0333$ in the full-time sample, $t_{\hat{y}} = 0.1583$ and $t_{\hat{u}} = 0.0191$ in male-dominated sectors, and $t_{\hat{y}} = 0.1649$ and $t_{\hat{u}} = 0.0357$ in female-dominated sectors. All statistics are significant at 1% level. The null hypothesis of similar distributions is strongly rejected, confirming that the distributions among men and women differ.

3.5 Conclusion

This work investigated the effects of gender sectoral segregation in the UK labour market by looking at average differences in contracts and wages for male and female workers over the period 2005 and 2020. Despite the national interest of the UK Government in promoting gender equality for all, evidence suggests that women and men are treated unfairly, especially in female-dominated sectors. This may potentially influence their choices in joining the labour force, their employment and career aspirations.

The disparity of opportunities in the labour market seems to be shaped by gender-based sectoral segregation where gender stereotypes still impact how employers perceive the skills of men and women. We indeed found that systematic contractual features that are preferred by women are quite common in female-dominated sectors dominated (even among men). Female-dominated sectors pay less in economic terms than male-dominated sectors for both men and women. The coefficient effect explains most of the gap rather than differences in accumulated human capital. However, women possess unobservable skills and behaviour that contribute to their positive selection at the bottom of the wage distribution, especially in female-dominated sectors, and make them less likely to be preferred to men.

This analysis has policy implications. Our findings can provide policy-makers with the empirical evidence in support of appropriate reforms in favour of vulnerable categories of workers (i.e., women, mothers, and immigrants) and policies designed to sustain long-run economic growth, especially in current times when the UK is facing new challenges (pandemic and the end of freedom of movement after Brexit). In line with Hyland et al.'s (2020) evidence, the institutional setting may affect women's access to the labour force and to entrepreneurial activities. Since women face more challenges than their male counterparts regarding labour participation, access to jobs and career opportunities, this gap could potentially widen in the post-pandemic.

Bibliography

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller (2010), "Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program." *Journal of the American statistical Association*, 105, 493–505.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller (2015), "Comparative politics and the synthetic control method." *American Journal of Political Science*, 59, 495–510.

Abadie, Alberto and Javier Gardeazabal (2003), "The economic costs of conflict: A case study of the basque country." *American economic review*, 93, 113–132.

Abadie, Alberto and Guido W Imbens (2006), "Large sample properties of matching estimators for average treatment effects." *econometrica*, 74, 235–267.

Acemoglu, Daron, Simon Johnson, James A Robinson, and Pierre Yared (2008), "Income and democracy." *American Economic Review*, 98, 808–42.

Alon, Titan M, Matthias Doepke, Jane Olmstead-Rumsey, and Michele Tertilt (2020), "The impact of covid-19 on gender equality." Technical report, National Bureau of economic research.

Aquaro, Michele and Pavel Čížek (2013), "One-step robust estimation of fixed-effects panel data models." *Computational Statistics & Data Analysis*, 57, 536–548.

Aquaro, Michele and Pavel Čížek (2014), "Robust estimation of dynamic fixed-effects panel data models." *Statistical Papers*, 55, 169–186.

Arellano, Manuel (1987), "Practitioners' corner: Computing robust standard errors for within-groups estimators." *Oxford bulletin of Economics and Statistics*, 49, 431–434.

Atkinson, Anthony Curtes (1985), "Plots, transformations and regression; an introduction to graphical methods of diagnostic regression analysis." Technical report.

Banerjee, Mousumi and Edward W Frees (1997), "Influence diagnostics for linear longitudinal models." *Journal of the American Statistical Association*, 92, 999–1005.

Belotti, Federico and Franco Peracchi (2020), "Fast leave-one-out methods for inference, model selection, and diagnostic checking." *The Stata Journal*, 20, 785–804, URL https://doi.org/10.1177/1536867X20976312.

Berka, Martin, Michael B. Devereux, and Charles Engel (2018), "Real exchange rates and sectoral productivity in the eurozone." *American Economic Review*, 108, 1543–81.

Bertoli, Simone, Vianney Dequiedt, and Yves Zenou (2016), "Can selective immigration policies reduce migrants' quality?" *Journal of Development Economics*, 119, 100–109.

Bertrand, Marianne (2011), "New perspectives on gender." In *Handbook of labor economics*, volume 4, 1543–1590, Elsevier.

Bertrand, Marianne (2020), "Gender in the twenty-first century." In *AEA Papers and proceedings*, volume 110, 1–24.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004), "How much should we trust differences-in-differences estimates?" *The Quarterly journal of economics*, 119, 249–275.

Bertrand, Marianne and Kevin F Hallock (2001), "The gender gap in top corporate jobs." *ILR Review*, 55, 3–21.

Bettio, Francesca, Alina Verashchagina, Ingrid Mairhuber, and Aleksandra Kanjuo-Mrčela (2009), *Gender segregation in the labour market: Root causes, implications and policy responses in the EU*. Publications Office of the European Union Luxembourg.

Blackburn, Robert M, Jennifer Jarman, and Janet Siltanen (1993), "The analysis of occupational gender segregation over time and place: considerations of measurement and some new evidence." *Work, Employment and Society*, 7, 335–362.

Blau, Francine D and Lawrence M Kahn (2017), "The gender wage gap: Extent, trends, and explanations." *Journal of economic literature*, 55, 789–865.

Blinder, Alan S (1973), "Wage discrimination: reduced form and structural estimates." *Journal of Human resources*, 436–455.

Bollen, Kenneth A and Robert W Jackman (1985), "Regression diagnostics: An expository treatment of outliers and influential cases." *Sociological Methods & Research*, 13, 510–542.

Booth, Alison L (2009), "Gender and competition." *Labour Economics*, 16, 599–606.

Borjas, George J (1980), "The relationship between wages and weekly hours of work: The role of division bias." *The Journal of Human Resources*, 15, 409–423.

Borjas, George J (1987), "Self-selection and the earnings of immigrants." *The American economic review*, 531–553.

Borjas, George J, Ilpo Kauppinen, and Panu Poutvaara (2019), "Self-selection of emigrants: Theory and evidence on stochastic dominance in observable and unobservable characteristics." *The Economic Journal*, 129, 143–171.

Boushey, Heather et al. (2005), "Are women opting out? debunking the myth." *Center for Economic and Policy Research, Briefing Paper*.

Bramati, Maria Caterina and Christophe Croux (2007), "Robust estimators for the fixed effects panel data model." *The econometrics journal*, 10, 521–540.

British Council (2016), *Gender Equality and Empowerment of Women and Girls in the UK*. URL https://www.britishcouncil.org/sites/default/files/gender_equality_and_empowerment_in_the_uk.pdf.

Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller (2011), "Robust inference with multiway clustering." *Journal of Business & Economic Statistics*, 29, 238–249.

Cameron, A Colin and Pravin K Trivedi (2005), *Microeconometrics: methods and applications*. Cambridge university press.

Campos-Soria, Juan Antonio and Miguel Angel Ropero-García (2016), "Gender segregation and earnings differences in the spanish labour market." *Applied Economics*, 48, 4143–4155.

Card, David (1992), "Using regional variation in wages to measure the effects of the federal minimum wage." *Ilr Review*, 46, 22–37.

Card, David, Lawrence F Katz, and Alan B Krueger (1994), "Comment on david neumark and william wascher,"employment effects of minimum and subminimum wages: Panel data on state minimum wage laws"." *ILR Review*, 47, 487–497.

Card, David and Alan B Krueger (2000), "Minimum wages and employment: a case study of the fast-food industry in new jersey and pennsylvania: reply." *American Economic Review*, 90, 1397–1420.

Cattaneo, Matias D, Michael Jansson, and Xinwei Ma (2019), "Two-step estimation and inference with possibly many included covariates." *The Review of Economic Studies*, 86, 1095–1122.

Cattaneo, Matias D, Michael Jansson, and Whitney K Newey (2018), "Inference in linear regression models with many covariates and heteroscedasticity." *Journal of the American Statistical Association*, 113, 1350–1361.

Chatterjee, Samprit and Ali S Hadi (1986), "Influential observations, high leverage points, and outliers in linear regression." *Statistical Science*, 379–393.

Chatterjee, Samprit and Ali S Hadi (1988), "Impact of simultaneous omission of a variable and an observation on a linear regression equation." *Computational Statistics & Data Analysis*, 6, 129–144.

Chesher, Andrew and Ian Jewitt (1987), "The bias of a heteroskedasticity consistent covariance matrix estimator." *Econometrica: Journal of the Econometric Society*, 1217–1222.

Chiswick, Barry (1999), "Are immigrants favorably self-selected?" *American Economic Review*, 89, 181–185.

Chiswick, Barry R (1978), "The effect of americanization on the earnings of foreign-born men." *Journal of political Economy*, 86, 897–921.

Chiswick, Barry R (1986), "Is the new immigration less skilled than the old?" *Journal of Labor Economics*, 4, 168–192.

Cook, R Dennis (1979), "Influential observations in linear regression." *Journal of the American Statistical Association*, 74, 169–174.

Cortes, Patricia and Jessica Pan (2018), "Occupation and gender." *The Oxford handbook of women and the economy*, 425–452.

Cortese, Charles F, R Frank Falk, and Jack K Cohen (1976), "Further considerations on the methodological analysis of segregation indices." *American sociological review*, 630–637.

Cribari-Neto, Francisco (2004), "Asymptotic inference under heteroskedasticity of unknown form." *Computational Statistics & Data Analysis*, 45, 215–233.

Cribari-Neto, Francisco and Wilton Bernardino da Silva (2011), "A new heteroskedasticity-consistent covariance matrix estimator for the linear regression model." *AStA Advances in Statistical Analysis*, 95, 129–146.

Cribari-Neto, Francisco, Tatiene C Souza, and Klaus LP Vasconcellos (2007), "Inference under heteroskedasticity and leveraged data." *Communications in Statistics - Theory and Methods*, 36, 1877–1888.

Czymara, Christian S, Alexander Langenkamp, and Tomás Cano (2020), "Cause for concerns: gender inequality in experiencing the covid-19 lockdown in germany." *European Societies*, 1–14.

Davidson, Russell, James G MacKinnon, et al. (1993), "Estimation and inference in econometrics." *OUP Catalogue*.

Doepke, Matthias and Michéle Tertilt (2016), "Families in macroeconomics." In *Handbook of macroeconomics*, volume 2, chapter 3, 1789–1891, Elsevier.

Donald, Stephen G. and G.S. Maddala (1993), "24 identifying outliers and influential observations in econometric models." In *Econometrics*, volume 11 of *Handbook of Statistics*, 663 – 701, Elsevier.

Efron, Bradley (1982), *The jackknife, the bootstrap, and other resampling plans*, volume 38. Siam.

Égert, Balázs (2016), "Regulation, institutions, and productivity: New macroeconomic evidence from oecd countries." *American Economic Review*, 106, 109–13.

Eicker, Friedhelm (1967), "Limit theorems for regressions with unequal and dependent errors." In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, 59–82.

Ellieroth, Kathrin et al. (2019), "Spousal insurance, precautionary labor supply, and the business cycle-a quantitative analysis." In *2019 Meeting Papers*, 1134, Society for Economic Dynamics.

Folke, Olle and Johanna Karin Rickne (2020), "Sexual harassment and gender inequality in the labor market."

Freedman, David A and Stephen C Peters (1984), "Bootstrapping an econometric model: Some empirical results." *Journal of Business & Economic Statistics*, 2, 150–158.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini (2003), "Performance in competitive environments: Gender differences." *The quarterly journal of economics*, 118, 1049–1074.

Gneezy, Uri and Aldo Rustichini (2004), "Gender and competition at a young age." *American Economic Review*, 94, 377–381.

Godfrey, LG (2006), "Tests for regression models with heteroskedasticity of unknown form." *Computational Statistics & Data Analysis*, 50, 2715–2733.

Goldin, Claudia (2006), "The quiet revolution that transformed women's employment, education, and family." *American economic review*, 96, 1–21.

Goldin, Claudia (2014), "A grand gender convergence: Its last chapter." *American Economic Review*, 104, 1091–1119.

Goldin, Claudia (2022), "Understanding the economic impact of covid-19 on women." Technical report, National Bureau of Economic Research.

Gomis, Roger, Steven Kapsos, and Ste Kuhn (2020), "World employment and social outlook: trends 2020." *Geneva: ILO*.

Gould, Eric D and Omer Moav (2016), "Does high inequality attract high skilled immigrants?" *The Economic Journal*, 126, 1055–1091.

Greene, William H (2012), *Econometric analysis*. Pearson Education.

Gupta, Alisha Haridasani (2020), "Why some women call this recession a "shecession,"." *New York Times*, 9.

Hansen, Bruce E (2019), "Econometrics." Unpublished manuscript. Latest version: February 2019.

Hansen, Christian B (2007), "Asymptotic properties of a robust variance matrix estimator for panel data when t is large." *Journal of Econometrics*, 141, 597–620.

Hayes, Andrew F and Li Cai (2007), "Using heteroskedasticity-consistent standard error estimators in ols regression: An introduction and software implementation." *Behavior research methods*, 39, 709–722.

Hendry, David F (1984), "Monte carlo experimentation in econometrics." *Handbook of econometrics*, 2, 937–976.

Hinkley, David V (1977), "Jackknifing in unbalanced situations." *Technometrics*, 19, 285–292.

Horn, Susan D, Roger A Horn, and David B Duncan (1975), "Estimating heteroscedastic variances in linear models." *Journal of the American Statistical Association*, 70, 380–385.

Hoynes, Hilary, Douglas L Miller, and Jessamyn Schaller (2012), "Who suffers during recessions?" *Journal of Economic perspectives*, 26, 27–48.

Huber, Peter J et al. (1967), "The behavior of maximum likelihood estimates under nonstandard conditions." In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, 221–233, University of California Press.

Huh, Myung-Hoe (1993), "A note on large-sample property of cook's distance." *Journal of the Japanese Society of Computational Statistics*, 6, 75–84.

Hyland, Marie, Simeon Djankov, and Pinelopi Koujianou Goldberg (2020), "Gendered laws and women in the workforce." *American Economic Review: Insights*, 2, 475–90.

James, David R and Karl E Taeuber (1985), "Measures of segregation." *Sociological methodology*, 15, 1–32.

Jann, Ben et al. (2008), "A stata implementation of the blinder-oaxaca decomposition." *Stata journal*, 8, 453–479.

Johnston, Ian (2021), "UK risks 'turning clock back' on gender equality in pandemic." URL https://www.theguardian.com/lifeandstyle/2021/feb/09/uk-risks-turning-clock-back-on-gender-equality-in-pandemic. [Online; posted 9-February-2021].

Kaur, Sanmeet (2020), *OECD Employment Outlook 2020*. URL https://www.fawcettsociety.org.uk/Handlers/Download.ashx?IDMF=bdb30c2d-7b79-4b02-af09-72d0e25545b5.

Kellogg, Maxwell, Magne Mogstad, Guillaume A Pouliot, and Alexander Torgovitsky (2021), "Combining matching and synthetic control to tradeoff biases from extrapolation and interpolation." *Journal of the American Statistical Association*, 116, 1804–1816.

Kezdi, Gabor (2003), "Robust standard error estimation in fixed-effects panel models." *Available at SSRN 596988*.

Kim, Myung Geun (2017), "A cautionary note on the use of cook's distance." *Communications for Statistical Applications and Methods*, 24, 317–324.

Kitagawa, Evelyn M (1955), "Components of a difference between two rates." *Journal of the american statistical association*, 50, 1168–1194.

Kiviet, Jan Frederik et al. (2012), *Monte Carlo simulation for econometricians*. now publishers.

Lawrance, AJ (1995), "Deletion influence and masking in regression." *Journal of the Royal Statistical Society: Series B (Methodological)*, 57, 181–189.

Long, J Scott and Laurie H Ervin (2000), "Using heteroscedasticity consistent standard errors in the linear regression model." *The American Statistician*, 54, 217–224.

MacKinnon, James G (2013), "Thirty years of heteroskedasticity-robust inference." In *Recent advances and future directions in causality, prediction, and specification analysis*, 437–461, Springer.

MacKinnon, James G and Halbert White (1985), "Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties." *Journal of econometrics*, 29, 305–325.

Martín, Nirian (2015), "Diagnostics in a simple correspondence analysis model: An approach based on cook's distance for log-linear models." *Journal of Multivariate Analysis*, 136, 175–189.

Martín, Nirian and Leandro Pardo (2009), "On the asymptotic distribution of cook's distance in logistic regression models." *Journal of Applied Statistics*, 36, 1119–1146.

Moir, Hazel and Joy Selby Smith (1979), "Industrial segregation in the australian labour market." *Journal of industrial Relations*, 21, 281–291.

Morchio, Iacopo and Christian Moser (2021), "The gender pay gap: Micro sources and macro consequences."

Mordaunt, Penny (2019), "Gender equality at every stage: a roadmap for change."

Mumford, Karen and Peter N Smith (2008), "What determines the part-time and gender earnings gaps in britain: evidence from the workplace." *Oxford Economic Papers*, 61, i56–i75.

Oaxaca, Ronald (1973), "Male-female wage differentials in urban labor markets." *International economic review*, 693–709.

OECD (2020), *2020 Sex and Power Index from the Fawcett Society*. URL https://www.oecd-ilibrary.org/content/publication/1686c758-en.

Olivetti, Claudia and Barbara Petrongolo (2014), "Gender gaps across countries and skills: Demand, supply and the industry structure." *Review of Economic Dynamics*, 17, 842–859.

Olivetti, Claudia and Barbara Petrongolo (2016), "The evolution of gender gaps in industrialized countries." *Annual review of Economics*, 8, 405–434.

Open Society Foundations (2020), "Q&a: How migrant women in the uk are joining together in the era of covid-19." URL http://web.archive.org/web/20080207010024/http://www.808multimedia.com/winnt/kernel.htm.

Parey, Matthias, Jens Ruhose, Fabian Waldinger, and Nicolai Netz (2017), "The selection of high-skilled emigrants." *Review of Economics and Statistics*, 99, 776–792.

Petrongolo, Barbara (2004), "Gender segregation in employment contracts." *Journal of the European Economic Association*, 2, 331–345.

Pinho, Luis Gustavo B, Juvêncio S Nobre, and Julio M Singer (2015), "Cook's distance for generalized linear mixed models." *Computational Statistics & Data Analysis*, 82, 126–136.

Poon, Wai-Yin and Yat Sun Poon (2001), "Conditional local influence in case-weights linear regression." *British Journal of Mathematical and Statistical Psychology*, 54, 177–191.

Reuben, Ernesto, Matthew Wiswall, and Basit Zafar (2017), "Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender." *The Economic Journal*, 127, 2153–2186.

Rosenbaum, Paul R and Donald B Rubin (1983), "The central role of the propensity score in observational studies for causal effects." *Biometrika*, 70, 41–55.

Rosenbaum, Paul R and Donald B Rubin (1985), "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score." *The American Statistician*, 39, 33–38.

Rousseeuw, Peter J (1991), "A diagnostic plot for regression outliers and leverage points." *Computational Statistics & Data Analysis*, 11, 127–129.

Rousseeuw, Peter J and Bert C Van Zomeren (1990), "Unmasking multivariate outliers and leverage points." *Journal of the American Statistical association*, 85, 633–639.

Saccardo, Silvia, Aniela Pietrasz, and Uri Gneezy (2018), "On the size of the gender difference in competitiveness." *Management Science*, 64, 1541–1554.

Scarborough, William J, Katherine Sobering, Ronald Kwon, and Mehr Mumtaz (2021), "The costs of occupational gender segregation in high-tech growth and productivity across us local labor markets." *Socio-Economic Review*.

Schularick, Moritz and Alan M Taylor (2012), "Credit booms gone bust: Monetary policy, leverage cycles, and financial crises, 1870-2008." *American Economic Review*, 102, 1029–61.

Silva, JMC Santos (2001), "Influence diagnostics and estimation algorithms for powell's scls." *Journal of Business & Economic Statistics*, 19, 55–62.

Şimşek, Esra and Mehmet Orhan (2016), "Heteroskedasticity-consistent covariance matrix estimators in small samples with high leverage points." *Theoretical Economics Letters*, 6, 658.

Stock, James H and Mark W Watson (2008), "Heteroskedasticity-robust standard errors for fixed effects panel data regression." *Econometrica*, 76, 155–174.

Verardi, Vincenzo and Christophe Croux (2009), "Robust regression in stata." *The Stata Journal*, 9, 439–453.

Verbeek, Marno (2008), *A guide to modern econometrics*. John Wiley & Sons.

Watts, Martin (1992), "How should occupational sex segregation be measured?" *Work, Employment and Society*, 6, 475–487.

Watts, Martin (1995), "Divergent trends in gender segregation by occupation in the united states: 1970–92." *Journal of Post Keynesian Economics*, 17, 357–378.

Watts, Martin (1998), "Occupational gender segregation: Index measurement and econometric modeling." *Demography*, 35, 489–496.

White, Halbert (1980), "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity." *Econometrica: Journal of the Econometric Society*, 817–838.

White, Halbert (1984), *Asymptotic theory for Econometricians*. Academic press.

Zoloth, Barbara S (1976), "Alternative measures of school segregation." *Land Economics*, 52, 278–298.

Appendix A

Mathematical Proofs

A.1 Leave-One-Out (L1O) Estimator

Following the L1O estimator for RE models in Banerjee and Frees (1997), we derive $\widehat{\boldsymbol{\beta}}_{(i)}$ using Woodbury's formula $(\mathrm{A}+\mathrm{BDC})^{-1}=\mathrm{A}^{-1}-\mathrm{A}^{-1}\mathrm{B}\left(\mathrm{D}^{-1}+\mathrm{CA}^{-1}\mathrm{B}\right)^{-1}\mathrm{CA}^{-1}$, where $\mathrm{A}=\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}$, $\mathrm{B}=-\widetilde{\mathbf{X}}_i'$, $\mathrm{C}=\widetilde{\mathbf{X}}_i$, and $\mathrm{D}=\mathbf{I}_T$.

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}_{(i)} &= \left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}-\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{X}}_i\right)^{-1}\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{Y}}-\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{y}}_i\right)\\
&=\left(\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}+\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}_i'\Big(\underbrace{\mathbf{I}_T-\widetilde{\mathbf{X}}_i\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}_i'}_{=\mathbf{I}_T-\mathbf{H}_i}\Big)^{-1}\widetilde{\mathbf{X}}_i\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\right)\times\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{Y}}-\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{y}}_i\right)\\
&=\underbrace{\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}'\widetilde{\mathbf{Y}}}_{=\widehat{\boldsymbol{\beta}}}-\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{y}}_i+\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}_i'(\mathbf{I}_T-\mathbf{H}_i)^{-1}\widetilde{\mathbf{X}}_i\underbrace{\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}'\widetilde{\mathbf{Y}}}_{=\widehat{\boldsymbol{\beta}}}\\
&\quad-\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}_i'(\mathbf{I}_T-\mathbf{H}_i)^{-1}\underbrace{\widetilde{\mathbf{X}}_i\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}_i'}_{=\mathbf{H}_i}\widetilde{\mathbf{y}}_i\\
&=\widehat{\boldsymbol{\beta}}-\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}_i'(\mathbf{I}_T-\mathbf{H}_i)^{-1}\left[(\mathbf{I}_T-\mathbf{H}_i)\widetilde{\mathbf{y}}_i-\widetilde{\mathbf{X}}_i\widehat{\boldsymbol{\beta}}+\mathbf{H}_i\widetilde{\mathbf{y}}_i\right]\\
&=\widehat{\boldsymbol{\beta}}-\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}_i'(\mathbf{I}_T-\mathbf{H}_i)^{-1}(\widetilde{\mathbf{y}}_i-\widetilde{\mathbf{X}}_i\widehat{\boldsymbol{\beta}})\\
&=\widehat{\boldsymbol{\beta}}-\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}_i'(\mathbf{I}_T-\mathbf{H}_i)^{-1}\widehat{\mathbf{u}}_i\\
&=\widehat{\boldsymbol{\beta}}-\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i, \qquad \text{(A.1)}
\end{aligned}
$$

where $\mathbf{M}_i^{-1} = (\mathbf{I}_T - \mathbf{H}_i)^{-1}$. Result (A.1) is the L1O estimator for FE in Belotti and Peracchi (2020). The sample mean of (A.1) is

$$\bar{\beta} \equiv \frac{1}{N}\sum_{i=1}^{N}\widehat{\beta}_{(i)} = \widehat{\beta} - \left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\frac{1}{N}\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i = \widehat{\beta} - \left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\mu^*, \tag{A.2}$$

where $\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{\beta} = N\widehat{\beta}$, $\mu^* = \frac{1}{N}\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i$ is a $k \times 1$ vector. Therefore, from (A.1) and (A.2) we get

$$\begin{aligned}\widehat{\beta}_{(i)} - \bar{\beta} &= \widehat{\beta} - \left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i - \widehat{\beta} + \left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\mu^* \\ &= -\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\left(\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i - \left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\mu^*\right)\end{aligned} \tag{A.3}$$

A.2 Derivation of the Jackknife Estimator

Following the procedure in Hansen (2019, pp. 324–326), the jackknife estimator of variance can be computed as

$$\widehat{\text{AVar}(\widehat{\beta})}_{jk} = \left(\frac{N-1}{N}\right)\sum_{i=1}^{N}\left(\widehat{\beta}_{(i)} - \bar{\beta}\right)\left(\widehat{\beta}_{(i)} - \bar{\beta}\right)' \tag{A.4}$$

$$= \left(\frac{N-1}{N}\right)\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\sum_{i=1}^{N}\left\{\left(\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i - \mu^*\right)\left(\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i - \mu^*\right)'\right\}\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1} \tag{A.5}$$

$$= \left(\frac{N-1}{N}\right)\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\left\{\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\mathbf{M}_i^{-1}\widetilde{\mathbf{X}}_i - N\frac{1}{N}\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i\mu^{*\prime}\right.$$

$$\left. - N\mu^*\frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{u}}_i'\mathbf{M}_i^{-1}\widetilde{\mathbf{X}}_i + N\mu^*\mu^{*\prime}\right\}\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1} \tag{A.6}$$

$$= \left(\frac{N-1}{N}\right)\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\left\{\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\mathbf{M}_i^{-1}\widetilde{\mathbf{X}}_i - N\mu^*\mu^{*\prime}\right\}\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1} \tag{A.7}$$

$$= \left(\frac{N-1}{N^2}\right)\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N}\right)^{-1}\left\{\widehat{\mathbf{V}}_N^3 - \mu^*\mu^{*\prime}\right\}\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N}\right)^{-1}, \tag{A.8}$$

where $\mu^* = \frac{1}{N}\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i$.

A.3 Proof Consistency of Transformed Residuals

We show that $\widehat{\mathbf{v}}_i = (\mathbf{I}_T - \mathbf{H}_i)^{-1}\widehat{\mathbf{u}}_i \xrightarrow{p} \widetilde{\mathbf{u}}_i$. We start from

$$\begin{aligned}\widehat{\mathbf{v}}_i - \widehat{\mathbf{u}}_i &= (\mathbf{I}_T - \mathbf{H}_i)^{-1}\widehat{\mathbf{u}}_i - \widehat{\mathbf{u}}_i \\ &= \big((\mathbf{I}_T - \mathbf{H}_i)^{-1} - \mathbf{I}_T\big)\widehat{\mathbf{u}}_i \\ &= \big((\mathbf{I}_T - \mathbf{H}_i)^{-1} - \mathbf{I}_T\big)\big(\widetilde{\mathbf{u}}_i - \widetilde{\mathbf{X}}_i(\widehat{\beta} - \beta)\big).\end{aligned} \tag{A.9}$$

Using Shwarz Inequality and Triangle Inequality for vectors and matrices (B.10 and B.13) in Hansen (2019, p.795), Equation (A.9) can be rewritten as

$$\begin{aligned}\|\widehat{\mathbf{v}}_i - \widehat{\mathbf{u}}_i\| &= \big\|\big((\mathbf{I}_T - \mathbf{H}_i)^{-1} - \mathbf{I}_T\big)\big(\widetilde{\mathbf{u}}_i - \widetilde{\mathbf{X}}_i(\widehat{\beta} - \beta)\big)\big\| \\ &\leq \big\|(\mathbf{I}_T - \mathbf{H}_i)^{-1} - \mathbf{I}_T\big\|\big\|\big(\widetilde{\mathbf{u}}_i - \widetilde{\mathbf{X}}_i(\widehat{\beta} - \beta)\big)\big\|\end{aligned} \tag{A.10}$$

Using Woodbury's formula $(\mathrm{A}+\mathrm{BDC})^{-1} = \mathrm{A}^{-1} - \mathrm{A}^{-1}\mathrm{B}(\mathrm{D}^{-1} + \mathrm{CA}^{-1}\mathrm{B})^{-1}\mathrm{CA}^{-1}$ with $\mathrm{A} = \mathbf{I}_T$, $\mathrm{B} = \widetilde{\mathbf{X}}_i$, $\mathrm{C} = \widetilde{\mathbf{X}}_i'$, and $\mathrm{D} = \widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}$, $(\mathbf{I}_T - \mathbf{H}_i)^{-1}$ can be rewritten as follows

$$(\mathbf{I}_T - \mathbf{H}_i)^{-1} = \big(\mathbf{I}_T - \widetilde{\mathbf{X}}_i(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}_i'\big)^{-1} = \mathbf{I}_T + \widetilde{\mathbf{X}}_i(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} - \widetilde{\mathbf{X}}_i'\widetilde{\mathbf{X}}_i)^{-1}\widetilde{\mathbf{X}}_i' \tag{A.11}$$

and, hence, the first component in (A.10)

$$\begin{aligned}\big\|(\mathbf{I}_T - \mathbf{H}_i)^{-1} - \mathbf{I}_T\big\| &= \|\widetilde{\mathbf{X}}_i(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} - \widetilde{\mathbf{X}}_i'\widetilde{\mathbf{X}}_i)^{-1}\widetilde{\mathbf{X}}_i'\| \\ &\leq \|\widetilde{\mathbf{X}}_i\|^2\big\|\big(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} - \widetilde{\mathbf{X}}_i'\widetilde{\mathbf{X}}_i\big)^{-1}\big\| \\ &= \frac{1}{N}\|\widetilde{\mathbf{X}}_i\|^2\left\|\left(\frac{1}{N}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} - \frac{1}{N}\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{X}}_i\right)^{-1}\right\|.\end{aligned} \tag{A.12}$$

Then, using expression (A.12) in inequality (A.10)

$$\|\widehat{\mathbf{v}}_i - \widehat{\mathbf{u}}_i\| \leq \frac{1}{N}\|\widetilde{\mathbf{X}}_i\|^2 \left\| \left(\frac{1}{N}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} - \frac{1}{N}\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{X}}_i\right)^{-1} \right\| \left(\|\widetilde{\mathbf{u}}_i\| + \|\widetilde{\mathbf{X}}_i\| \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|\right) \tag{A.13}$$
$$= o_p(1)$$

where the first component of (A.13) is $O_p(N^{-1})$ under ASM.4.i for $r \geq 2$; the second component involves $\mathbf{S}_{XX} + o_p(1)$ as $N^{-1}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} \xrightarrow{p} \mathbb{E}(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}) = \mathbf{S}_{XX} > 0$ by the Central Limit Theorem; the last component is $O_p(1)$ because the random variables in parenthesis are $O_p(1)$ under ASM.4.i-ii, and $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\| \xrightarrow{p} 0$ is $o_p(1)$. Therefore, the overall expression is bounded above by an $o_p(1)$ random variable. Note that

$$\|\widehat{\mathbf{u}}_i - \widetilde{\mathbf{u}}_i\| \leq \|\widetilde{\mathbf{X}}_i\| \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\| \tag{A.14}$$

because $\widetilde{\mathbf{X}}_i$ is $O_p(1)$ by ASM.4.i and $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\| \xrightarrow{p} 0$, $\|\widehat{\mathbf{u}}_i - \widetilde{\mathbf{u}}_i\|$ is $o_p(1)$. Therefore, $\widehat{\mathbf{u}}_i \xrightarrow{p} \widetilde{\mathbf{u}}_i$ as $N \to \infty$ and T fixed. Using result in (A.13) and (A.14), we obtain the desired result

$$\widehat{\mathbf{v}}_i = \widehat{\mathbf{u}}_i + o_p(1) \xrightarrow{p} \widetilde{\mathbf{u}}_i \quad \text{as } N \to \infty \text{ and } T \text{ fixed.} \tag{A.15}$$

Result (A.15) shows that the transformed standard errors $\widehat{\mathbf{v}}_i$ are a uniformly consistent estimator for the error term $\widetilde{\mathbf{u}}_i$. This result guarantees the consistency of any other formula of alternative HC estimators, such as *PHC*3 and *PHC*6. In fact, using Equation (A.15) we can show that

$$\widehat{\widehat{\mathbf{V}}}_N - \widehat{\mathbf{V}}_N = \frac{1}{N}\sum_{i=1}^{N} \widetilde{\mathbf{X}}_i'\widehat{\mathbf{v}}_i\widehat{\mathbf{v}}_i'\widetilde{\mathbf{X}}_i - \frac{1}{N}\sum_{i=1}^{N} \widetilde{\mathbf{X}}_i'\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\widetilde{\mathbf{X}}_i$$
$$= \frac{1}{N}\sum_{i=1}^{N} \widetilde{\mathbf{X}}_i'\Big(\widehat{\mathbf{v}}_i(\widehat{\mathbf{v}}_i' - \widehat{\mathbf{u}}_i')(\widehat{\mathbf{v}}_i + \widehat{\mathbf{u}}_i)\widehat{\mathbf{u}}_i'\Big)\widetilde{\mathbf{X}}_i \tag{A.16}$$

Then,

$$\begin{aligned}
\left\|\widehat{\widehat{\mathbf{V}}}_N - \widehat{\mathbf{V}}_N\right\| &\leq \frac{1}{N}\sum_{i=1}^{N}\|\widetilde{\mathbf{X}}_i'\widehat{\mathbf{v}}_i(\widehat{\mathbf{v}}_i' - \widehat{\mathbf{u}}_i')\widetilde{\mathbf{X}}_i\| + \frac{1}{N}\sum_{i=1}^{N}\|\widetilde{\mathbf{X}}_i'(\widehat{\mathbf{v}}_i + \widehat{\mathbf{u}}_i)\widehat{\mathbf{u}}_i'\widetilde{\mathbf{X}}_i\| \\
&\leq \max_{1\leq i\leq N}\|\widehat{\mathbf{v}}_i - \widehat{\mathbf{u}}_i\|\left(\frac{1}{N}\sum_{i=1}^{N}\|\widetilde{\mathbf{X}}_i\widetilde{\mathbf{X}}_i'\widehat{\mathbf{v}}_i\| + \frac{1}{N}\sum_{i=1}^{N}\|\widetilde{\mathbf{X}}_i\widetilde{\mathbf{X}}_i'\widehat{\mathbf{u}}_i'\|\right) \\
&\leq \max_{1\leq i\leq N}\|\widehat{\mathbf{v}}_i - \widehat{\mathbf{u}}_i\|\left(\frac{1}{N}\sum_{i=1}^{N}\|\widetilde{\mathbf{X}}_i\|^2\|\widehat{\mathbf{v}}_i\| + \frac{1}{N}\sum_{i=1}^{N}\|\widetilde{\mathbf{X}}_i\|^2\|\widehat{\mathbf{u}}_i\|\right) \qquad \text{(A.17)} \\
&= o_p(1)
\end{aligned}$$

where the first term of (A.17) is $o_p(1)$ from result (A.13); the two components in parenthesis are the sums of random variables with finite means both converging in probability to $\mathbb{E}\left(\|\widetilde{\mathbf{X}}_i\|^2\|\widetilde{\mathbf{u}}_i\|\right)$ by ASM.4 and results (A.14)-(A.15) and, therefore, $O_p(1)$. Their product is $o_p(1)$.

The last step left to show is $\widehat{\mathbf{V}}_N \xrightarrow{p} \mathbf{V}$ such that $\widehat{\widehat{\mathbf{V}}}_N = \widehat{\mathbf{V}}_N + o_p(1) \xrightarrow{p} \mathbf{V}$ as $N \to \infty$ and T fixed. Following Hansen (2019, pp.230–232), we start from the definition of conventional robust variance-covariance matrix

$$\begin{aligned}
\widehat{\mathbf{V}}_N &= \frac{1}{N}\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\widetilde{\mathbf{X}}_i \\
&\frac{1}{N}\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{u}}_i\widetilde{\mathbf{u}}_i'\widetilde{\mathbf{X}}_i + \frac{1}{N}\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'(\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i' - \widetilde{\mathbf{u}}_i\widetilde{\mathbf{u}}_i')\widetilde{\mathbf{X}}_i \qquad \text{(A.18)}
\end{aligned}$$

where the first component of Equation (A.18) converges in probability to $\mathbb{E}(\widetilde{\mathbf{X}}_i'\Sigma_i\widetilde{\mathbf{X}}_i) = \mathbf{V}_i$ by ASM.2.ii by LIE and ASM.1 with finite limit $\mathbf{V}$ under THM 6.16 in Hansen (2019, p.189) for sequences of *inid* random variables, provided that ASM.5 and ASM.4.i hold. The second component needs to converge in probability to zero to claim consistency of $\widehat{\mathbf{V}}_N$. Applying matrix norm to (A.18), we get

$$\left\|\frac{1}{N}\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'(\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'-\widetilde{\mathbf{u}}_i\widetilde{\mathbf{u}}_i')\widetilde{\mathbf{X}}_i\right\| \leq \frac{1}{N}\sum_{i=1}^{N}\left\|\widetilde{\mathbf{X}}_i'(\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'-\widetilde{\mathbf{u}}_i\widetilde{\mathbf{u}}_i')\widetilde{\mathbf{X}}_i\right\| \leq \frac{1}{N}\sum_{i=1}^{N}\|\widetilde{\mathbf{X}}_i\|^2\|\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'-\widetilde{\mathbf{u}}_i\widetilde{\mathbf{u}}_i'\| \tag{A.19}$$

Note that

$$\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i' = \left(\widetilde{\mathbf{u}}_i-\widetilde{\mathbf{X}}_i(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\right)\left(\widetilde{\mathbf{u}}_i-\widetilde{\mathbf{X}}_i(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\right)' = \widetilde{\mathbf{u}}_i\widetilde{\mathbf{u}}_i'-\widetilde{\mathbf{u}}_i(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\widetilde{\mathbf{X}}_i'-\widetilde{\mathbf{X}}_i(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\widetilde{\mathbf{u}}_i'+\widetilde{\mathbf{X}}_i(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\widetilde{\mathbf{X}}_i' \tag{A.20}$$

Rearranging last line of Equation (A.20) and using the Triangle Inequality (B.14) and Schwarz Inequality (B.13), we obtain

$$\|\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'-\widetilde{\mathbf{u}}_i\widetilde{\mathbf{u}}_i'\| \leq 2\left\|\widetilde{\mathbf{X}}_i(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\widetilde{\mathbf{u}}_i'\right\|+\|\widetilde{\mathbf{X}}_i\|^2\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|^2 \leq 2\|\widetilde{\mathbf{X}}_i\|\|\widetilde{\mathbf{u}}_i\|\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|+\|\widetilde{\mathbf{X}}_i\|^2\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|^2 \tag{A.21}$$

Plugging (A.21) in (A.19)

$$\begin{aligned}\left\|\frac{1}{N}\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'(\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'-\widetilde{\mathbf{u}}_i\widetilde{\mathbf{u}}_i')\widetilde{\mathbf{X}}_i\right\| &\leq \frac{1}{N}\sum_{i=1}^{N}\|\widetilde{\mathbf{X}}_i\|^2\|\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'-\widetilde{\mathbf{u}}_i\widetilde{\mathbf{u}}_i'\| \\ &\leq \frac{1}{N}\sum_{i=1}^{N}\|\widetilde{\mathbf{X}}_i\|^2\left\{2\|\widetilde{\mathbf{X}}_i\|\|\widetilde{\mathbf{u}}_i\|\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|+\|\widetilde{\mathbf{X}}_i\|^2\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|^2\right\} \\ &\leq 2\left(\frac{1}{N}\sum_{i=1}^{N}\|\widetilde{\mathbf{X}}_i\|^3\|\widetilde{\mathbf{u}}_i\|\right)\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|+\left(\frac{1}{N}\sum_{i=1}^{N}\|\widetilde{\mathbf{X}}_i\|^4\right)\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|^2 \\ &= o_p(1)\end{aligned} \tag{A.22}$$

where the average in the first parenthesis is $O_p(1)$ because it is the mean of random variables bounded above by finite quantities, that is, $\mathbb{E}\left(\left\|\widetilde{\mathbf{X}}_i\right\|^3\left\|\widetilde{\mathbf{u}}_i\right\|\right) \leq \left(\mathbb{E}\left\|\widetilde{\mathbf{X}}_i\right\|^3\right)^{\frac{3}{4}}\left(\mathbb{E}\left\|\widetilde{\mathbf{u}}_i\right\|^4\right)^{\frac{1}{4}}$ by evoking Hölder's Inequality (B.28) in Hansen (2019, p. 796) and under ASM.1 and ASM.4.i-ii; the average in the second parenthesis is $O_p(1)$ as the average of a random variable with finite mean by ASM.4.i; $\left\|\widehat{\beta}-\beta\right\| \xrightarrow{p} 0$ and, thus, is $o_p(1)$. It follows that $\widehat{\mathbf{V}}_N \xrightarrow{p} \mathbf{V}$ and, therefore, the desired result

$$\widehat{\widehat{\mathbf{V}}}_N = \widehat{\mathbf{V}}_N + o_p(1) \to \mathbf{V}. \tag{A.23}$$

A.4 Consistency of PHC6

The proposed estimator, PHC6, of the asymptotic variance-covariance matrix is

$$\widehat{\text{AVar}(\widehat{\beta})}_6 = c_6 \mathbf{S}_N^{-1} \widehat{\mathbf{V}}_N^6 \mathbf{S}_N^{-1}, \tag{A.24}$$

where the variance-covariance matrix is $\widehat{\mathbf{V}}_N^6 = \frac{1}{N}\sum_{i=1}^N \widetilde{\mathbf{X}}_i' \widehat{\mathbf{v}}_i \widehat{\mathbf{v}}_i' \widetilde{\mathbf{X}}_i$, and the matrix $\mathbf{M}_i$ has functional form

$$\mathbf{M}_i = \begin{cases} \mathbf{I}_T & \text{if } h_i^* < 2 \\ \mathbf{I}_T - \mathbf{H}_i & \text{otherwise} \end{cases} \tag{A.25}$$

where $h_i^* = max\{h_{i11}/\overline{h}_{11}, \ldots, h_{iTT}/\overline{h}_{TT}\}$ is the maximal individual leverage of unit i; and $\overline{h}_{tt} = N^{-1}\sum_{i=i}^N h_{itt}$ is the average leverage at time t, with h_{itt} being the individual leverage of unit i at time t. The finite sample correction of PHC6 is

$$c_6 = \begin{cases} \frac{(NT-1)N}{(NT-k)(N-1)} & \text{if } h_i^* < 2 \\ \frac{N-1}{N} & \text{otherwise} \end{cases}$$

As $N \to \infty$ and T is fixed, $\mathbf{M} = (\mathbf{I}_T - \mathbf{H}_i)$ because the selection criterion is $2 < \infty$. As argued above, $\mathbf{H}_i \xrightarrow{p} \mathbf{0}$ as $N \to \infty$ and T fixed because leverage measures are asymptotically negligible (Hansen, 2019, p.249). Therefore, PHC6 collapses to PHC3 that converges to PHC0 which is a consistent estimator of the asymptotic variance (Hansen, 2019, Theorem 7.7, p.232). From White's (1980) general result and under the above model assumptions and THM 7.7 in Hansen (2019, p.232), it follows $\widetilde{\Sigma} = \widehat{\Sigma} + o_p(1) \to \Sigma$ as $N \to \infty$ and T fixed such that $\widehat{\text{AVar}(\widehat{\beta})}_6 \xrightarrow{p} \text{AVar}(\widehat{\beta})$, making PHC6 consistent estimator of the sampling variance.

A.5 Derivation of the Distribution of W

The error term u_{it} is intrinsically heteroskedastic but not on average due to the presence of the scaling factor $z(\gamma)$. Let $\text{W} = \beta_0 + \sum_{j=1}^{J} \beta_j x_{it,j}$ with $\{x_{it,j}\}_{j=1}^{J}$. When $\gamma = 1$, the mean and variance of a random variable W with an unknown distribution are as follows

$$\mathbb{E}(\text{W}) = \mathbb{E}\left[\beta_0 + \sum_{j=1}^{J} \beta_j x_{it,j}\right] = \beta_0 + \sum_{j=1}^{J} \beta_j \mathbb{E}\left(x_{it,j}\right) = \beta_0 + \sum_{j=1}^{J} \beta_j \mu_{x_j} \tag{A.26}$$

$$\begin{aligned} \text{Var}\left(\text{W}\right) &= \text{Var}\left[\beta_0 + \sum_{j=1}^{J} \beta_j x_{it,j}\right] = \sum_{j=1}^{J} \beta_j^2 \,\text{Var}\left(x_{it,j}\right) + 2 \sum_{\substack{j,k=1 \\ j \neq k}}^{J} \beta_j \beta_k \text{Cov}(x_{it,j}, x_{it,k}) \\ &= \sum_{j=1}^{J} \beta_j^2 \sigma_{x_j}^2, \end{aligned} \tag{A.27}$$

where $\mathbb{E}\left(x_{it,j}\right) = \mu_{x_j}$, $\text{Var}\left(x_{it,j}\right) = \sigma_{x_j}^2$, and $\text{Cov}(x_{it,j}, x_{it,k}) = 0$ because the independence assumption guarantees that $\mathbb{E}\left(x_{it,j}, x_{it,k}\right) = \mathbb{E}\left(x_{it,j}\right)\mathbb{E}\left(x_{it,k}\right)$. The results are valid under independent and identically distributed (*iid*) random variables. By the assumptions of *iid* and normality of $\mathbf{x}_{it}$, the random variable W is normally distributed with mean (A.26) and variance (A.27). When the regressors are drawn from a standard normal distribution, (A.26) and (A.27) reduce to β_0 and $\sum_{j=1}^{J} \beta_j^2$, respectively. Thus, $\text{W} \sim \mathcal{N}\left(\beta_0, \sum_{j=1}^{J} \beta_j^2\right)$. Standardising W, we get $\frac{\text{W} - \mu_w}{\sigma_w^2} \sim \mathcal{N}\left(0, 1\right)$.

When $\gamma = 2$, the mean and variance of W are as follows

$$\begin{aligned}
\mathbb{E}(\mathrm{W}^2) &= \mathbb{E}\left[\left(\beta_0 + \sum_{j=1}^{J} \beta_j x_{it,j}\right)^2\right] \\
&= \beta_0^2 + \sum_{j=1}^{J} \beta_j^2\, \mathbb{E}\big(x_{it,j}^2\big) + 2\beta_0 \sum_{j=1}^{J} \beta_j\, \mathbb{E}\big(x_{it,j}\big) + 2 \sum_{\substack{j,k=1\\ j\neq k}}^{J} \beta_j\beta_k \mathbb{E}\big(x_{it,j}\big)\mathbb{E}\big(x_{it,j}\big) \qquad \text{(A.28)} \\
&= \beta_0^2 + \sum_{j=1}^{J} \beta_j^2 (\sigma_{x_j}^2 + \mu_{x_j}^2) + 2\beta_0 \sum_{j=1}^{J} \beta_j \mu_{x_j} + 2 \sum_{\substack{j,k=1\\ j\neq k}}^{J} \beta_j\beta_k \mu_{x_j}\mu_{x_j}
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}\big(\mathrm{W}^2\big) &= \mathrm{Var}\left[\left(\beta_0 + \sum_{j=1}^{J} \beta_j x_{it,j}\right)^2\right] \\
&= \sum_{j=1}^{J} \beta_k^4\, \mathrm{Var}\big(x_{it,j}^2\big) + 4\beta_0^2 \sum_{j=1}^{J} \beta_j^2\, \mathrm{Var}\big(x_{it,j}\big) + 4 \sum_{\substack{j,k=1\\ j\neq k}}^{J} \beta_j^2\beta_k^2 \mathrm{Var}\big(x_{it,j}, x_{it,j}\big) \\
&= \sum_{j=1}^{J} \beta_k^4 \sigma_{x_j}^2 + 4\beta_0^2 \sum_{j=1}^{J} \beta_j^2 \sigma_{x_j}^2 + 4 \sum_{\substack{j,k=1\\ j\neq k}}^{J} \beta_j^2\beta_k^2 \big(\sigma_{x_j}^2\sigma_{x_j}^2 + \sigma_{x_j}^2\mu_{x_j}^2 + \sigma_{x_j}^2\mu_{x_j}^2\big) \qquad \text{(A.29)}
\end{aligned}$$

where $\mathbb{E}\big(x_{it,j}^2\big) = \sigma_{x_j}^2 + \mu_{x_j}^2$, $\mathbb{E}\big(x_{it,j}, x_{it,k}\big) = \mathbb{E}\big(x_{it,j}\big)\mathbb{E}\big(x_{it,k}\big) = \mu_{x_j}\mu_{x_k}$, and $\mathrm{Var}\big(x_{it,j}, x_{it,k}\big) = \big[\mathbb{E}\big(x_{it,j}, x_{it,ks}\big)\big]^2 - \mathbb{E}\big(x_{it,j}\big)^2\mathbb{E}\big(x_{it,ks}\big)^2 = \sigma_{x_j}^2\sigma_{x_k s}^2 + \sigma_{x_j}^2\mu_{x_k}^2 + \sigma_{x_j}^2\mu_{x_k}^2$ because of the *iid* assumption. Any covariance among variables is null because of the assumption of independence. With standardised W, $\left(\frac{\mathrm{W}-\mu_w}{\sigma_w^2}\right)^2 \sim \chi_1^2$.

A.6 Leave-Two-Out (L2O) Estimator

Let $\widehat{\boldsymbol{\beta}}_{(i,j)}$ be the FE estimator without units i and j computed as follows

$$\widehat{\boldsymbol{\beta}}_{(i,j)} = \left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} - \widetilde{\mathbf{X}}_i'\widetilde{\mathbf{X}}_i - \widetilde{\mathbf{X}}_j'\widetilde{\mathbf{X}}_j\right)^{-1}\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}_i'\widetilde{\mathbf{y}}_i - \widetilde{\mathbf{X}}_j'\widetilde{\mathbf{y}}_j\right), \qquad \text{(A.30)}$$

where $\widetilde{\mathbf{X}} \in \mathbb{R}^{NT\times k}$, $\widetilde{\mathbf{Y}} \in \mathbb{R}^{NT\times 1}$, $\widetilde{\mathbf{X}}_i \in \mathbb{R}^{T_i\times k}$, $\widetilde{\mathbf{X}}_j \in \mathbb{R}^{T_j\times k}$, $\widetilde{\mathbf{y}}_i \in \mathbb{R}^{T_i\times 1}$, and $\widetilde{\mathbf{y}}_j \in \mathbb{R}^{T_j\times 1}$. Combining the matrices as follows $\widetilde{\mathbf{X}}_{ij} = \big[\widetilde{\mathbf{X}}_i' \;\; \widetilde{\mathbf{X}}_j'\big]' \in \mathbb{R}^{(T_i+T_j)\times k}$, $\widetilde{\mathbf{X}}_{ij}' = \big[\widetilde{\mathbf{X}}_i' \;\; \widetilde{\mathbf{X}}_j'\big] \in \mathbb{R}^{k\times(T_i+T_j)}$, and $\widetilde{\mathbf{y}}_{ij} =$

$\left[\widetilde{\mathbf{y}}_i \;\; \widetilde{\mathbf{y}}_j\right]' \in \mathbb{R}^{(T_i+T_j)\times 1}$, the standard Woodbury's formula, $(\mathrm{A}+\mathrm{BDC})^{-1} = \mathrm{A}^{-1} - \mathrm{A}^{-1}\mathrm{B}\left(\mathrm{D}^{-1} + \mathrm{CA}^{-1}\mathrm{B}\right)^{-1}\mathrm{CA}^{-1}$, can be applied where $\mathrm{A} = \widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}$, $\mathrm{B} = -\widetilde{\mathbf{X}}'_{ij}$, $\mathrm{C} = \widetilde{\mathbf{X}}_{ij}$, and $\mathrm{D} = \begin{bmatrix} \mathbf{I}_i & \mathbf{0}_{ij} \\ \mathbf{0}_{ji} & \mathbf{I}_j \end{bmatrix} = \mathbf{I}_i \oplus \mathbf{I}_j \equiv \mathbf{I}_{ij} \in \mathbb{R}^{(T_i+T_j)\times(T_i+T_j)}$, whose block diagonal identity matrices $\mathbf{I}_i$ and $\mathbf{I}_j$ are respectively of order $T_i \times T_i$ and $T_j \times T_j$, and the off-diagonal block matrices of zeros denoted with subscripts ij and ji are $T_i \times T_j$ and $T_j \times T_i$, respectively. With balanced panels, the notation simplifies to $T_i = T_j = T$.

Then, Equation (A.30) becomes

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{(i,j)} &= \left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} - \widetilde{\mathbf{X}}'_{ij}\widetilde{\mathbf{X}}_{ij}\right)^{-1}\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}'_{ij}\widetilde{\mathbf{y}}_{ij}\right) \\
&= \left(\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1} + \left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}'_{ij}\Big(\underbrace{\mathbf{I}_{ij} - \widetilde{\mathbf{X}}_{ij}\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}'_{ij}}_{=\mathbf{I}_{ij}-\overline{\mathbf{H}}_{ij}}\Big)^{-1}\widetilde{\mathbf{X}}_{ij}\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\right) \times \left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}'_{ij}\widetilde{\mathbf{y}}_{ij}\right) \\
&= \underbrace{\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}'\widetilde{\mathbf{Y}}}_{=\widehat{\boldsymbol{\beta}}} - \left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}'_{ij}\widetilde{\mathbf{y}}_{ij} + \left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}'_{ij}(\mathbf{I}_{ij} - \overline{\mathbf{H}}_{ij})^{-1}\widetilde{\mathbf{X}}_{ij}\underbrace{\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}'\widetilde{\mathbf{Y}}}_{=\widehat{\boldsymbol{\beta}}} \\
&\quad - \left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}'_{ij}(\mathbf{I}_{ij} - \overline{\mathbf{H}}_{ij})^{-1}\underbrace{\widetilde{\mathbf{X}}_{ij}\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}'_{ij}}_{=\overline{\mathbf{H}}_{ij}}\widetilde{\mathbf{y}}_{ij} \\
&= \widehat{\boldsymbol{\beta}} - \left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}'_{ij}(\mathbf{I}_{ij} - \overline{\mathbf{H}}_{ij})^{-1}\left[(\mathbf{I}_{ij} - \overline{\mathbf{H}}_{ij})\widetilde{\mathbf{y}}_{ij} - \widetilde{\mathbf{X}}_{ij}\widehat{\boldsymbol{\beta}} + \overline{\mathbf{H}}_{ij}\widetilde{\mathbf{y}}_{ij}\right] \\
&= \widehat{\boldsymbol{\beta}} - \left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}'_{ij}(\mathbf{I}_{ij} - \overline{\mathbf{H}}_{ij})^{-1}(\widetilde{\mathbf{y}}_{ij} - \widetilde{\mathbf{X}}_{ij}\widehat{\boldsymbol{\beta}}) \\
&= \widehat{\boldsymbol{\beta}} - \left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}'_{ij}(\mathbf{I}_{ij} - \overline{\mathbf{H}}_{ij})^{-1}\widehat{\mathbf{u}}_{ij} \\
&= \widehat{\boldsymbol{\beta}} - \left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}'_{ij}\overline{\mathbf{M}}_{ij}^{-1}\widehat{\mathbf{u}}_{ij}, \qquad \text{(A.31)}
\end{aligned}$$

where $\overline{\mathbf{M}}_{ij} = (\mathbf{I}_{ij} - \overline{\mathbf{H}}_{ij})$, with $\overline{\mathbf{M}}_{ij} = \begin{bmatrix} \mathbf{M}_i & -\mathbf{H}_{ij} \\ -\mathbf{H}'_{ij} & \mathbf{M}_j \end{bmatrix} = \begin{bmatrix} \mathbf{I}_i - \mathbf{H}_i & -\mathbf{H}_{ij} \\ -\mathbf{H}'_{ij} & \mathbf{I}_j - \mathbf{H}_j \end{bmatrix}$ and the leverage matrix $\overline{\mathbf{H}}_{ij} = \begin{bmatrix} \mathbf{H}_i & \mathbf{H}_{ij} \\ \mathbf{H}'_{ij} & \mathbf{H}_j \end{bmatrix} = \begin{bmatrix} \widetilde{\mathbf{X}}_i(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}'_i & \widetilde{\mathbf{X}}_i(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}'_j \\ \widetilde{\mathbf{X}}_j(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}'_i & \widetilde{\mathbf{X}}_j(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}'_j \end{bmatrix}$

Note that $\mathbf{H}_{ji} = \mathbf{H}'_{ij}$. Expanding last block of Equation (A.31) and using the formulae for the inverse of partitioned matrices[1], we have

$$\widetilde{\mathbf{X}}'_{ij}\overline{\mathbf{M}}_{ij}^{-1}\widehat{\mathbf{u}}_{ij} =$$

$$= \begin{bmatrix} \widetilde{\mathbf{X}}'_i & \widetilde{\mathbf{X}}'_j \end{bmatrix} \begin{bmatrix} \mathbf{M}_i & -\mathbf{H}_{ij} \\ -\mathbf{H}'_{ij} & \mathbf{M}_j \end{bmatrix}^{-1} \begin{bmatrix} \widehat{\mathbf{u}}_i \\ \widehat{\mathbf{u}}_j \end{bmatrix}$$

$$= \begin{bmatrix} \widetilde{\mathbf{X}}'_i & \widetilde{\mathbf{X}}'_j \end{bmatrix} \begin{bmatrix} \mathbf{M}_i^{-1} + \mathbf{M}_i^{-1}\mathbf{H}_{ij}(\mathbf{M}_j - \mathbf{H}'_{ij}\mathbf{M}_i^{-1}\mathbf{H}_{ij})^{-1}\mathbf{H}'_{ij}\mathbf{M}_i^{-1} & \mathbf{M}_i^{-1}\mathbf{H}_{ij}(\mathbf{M}_j - \mathbf{H}'_{ij}\mathbf{M}_i^{-1}\mathbf{H}_{ij})^{-1} \\ (\mathbf{M}_j - \mathbf{H}'_{ij}\mathbf{M}_i^{-1}\mathbf{H}_{ij})^{-1}\mathbf{H}'_{ij}\mathbf{M}_i^{-1} & (\mathbf{M}_j - \mathbf{H}'_{ij}\mathbf{M}_i^{-1}\mathbf{H}_{ij})^{-1} \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{u}}_i \\ \widehat{\mathbf{u}}_j \end{bmatrix}$$

$$= \begin{bmatrix} \left[\widetilde{\mathbf{X}}'_i\mathbf{M}_i^{-1} + \left(\widetilde{\mathbf{X}}'_i\mathbf{M}_i^{-1}\mathbf{H}_{ij} + \widetilde{\mathbf{X}}'_j\right)(\mathbf{M}_j - \mathbf{H}'_{ij}\mathbf{M}_i^{-1}\mathbf{H}_{ij})^{-1}\mathbf{H}'_{ij}\mathbf{M}_i^{-1}\right]' \\ \left[\left(\widetilde{\mathbf{X}}'_i\mathbf{M}_i^{-1}\mathbf{H}_{ij} + \widetilde{\mathbf{X}}'_j\right)(\mathbf{M}_j - \mathbf{H}'_{ij}\mathbf{M}_i^{-1}\mathbf{H}_{ij})^{-1}\right]' \end{bmatrix}' \begin{bmatrix} \widehat{\mathbf{u}}_i \\ \widehat{\mathbf{u}}_j \end{bmatrix}$$

$$= \widetilde{\mathbf{X}}'_i\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i + \left(\widetilde{\mathbf{X}}'_i\mathbf{M}_i^{-1}\mathbf{H}_{ij} + \widetilde{\mathbf{X}}'_j\right)\left(\mathbf{M}_j - \mathbf{H}'_{ij}\mathbf{M}_i^{-1}\mathbf{H}_{ij}\right)^{-1}\left(\mathbf{H}'_{ij}\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i + \widehat{\mathbf{u}}_j\right), \tag{A.32}$$

provided that $\mathbf{M}_i$ and $\left(\mathbf{M}_j - \mathbf{H}'_{ij}\mathbf{M}_i^{-1}\mathbf{H}_{ij}\right)$ are non-singular to be invertible.

Therefore, Equation (A.31) can be rewritten as follows

$$\begin{aligned} \widehat{\beta}_{(i,j)} &= \widehat{\beta} - \left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\left(\widetilde{\mathbf{X}}'_i\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i + \left(\widetilde{\mathbf{X}}'_i\mathbf{M}_i^{-1}\mathbf{H}_{ij} + \widetilde{\mathbf{X}}'_j\right)\left(\mathbf{M}_j - \mathbf{H}'_{ij}\mathbf{M}_i^{-1}\mathbf{H}_{ij}\right)^{-1}\left(\mathbf{H}'_{ij}\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i + \widehat{\mathbf{u}}_j\right)\right) \\ &= \widehat{\beta}_{(i)} - \left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\left(\widetilde{\mathbf{X}}'_i\mathbf{M}_i^{-1}\mathbf{H}_{ij} + \widetilde{\mathbf{X}}'_j\right)\left(\mathbf{M}_j - \mathbf{H}'_{ij}\mathbf{M}_i^{-1}\mathbf{H}_{ij}\right)^{-1}\left(\mathbf{H}'_{ij}\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i + \widehat{\mathbf{u}}_j\right) \end{aligned} \tag{A.33}$$

where $\widehat{\beta}_{(i)} = \widehat{\beta} - \left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}'_i\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i$.

[1] The formula used is

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B(D-CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D-CA^{-1}B)^{-1} \\ -(D-CA^{-1}B)^{-1}CA^{-1} & (D-CA^{-1}B)^{-1} \end{bmatrix}$$

provided that A^{-1} exists and $(D-CA^{-1}B)$ is invertible.

A.7 Distribution of Diagnostic Measures

A.7.1 Joint Influence

For simplicity, consider Equation (2.41) the case when $i = j$. We first establish the asymptotic properties of $\widehat{\boldsymbol{\beta}}_{(i)}$ and show its equivalence to $\widehat{\boldsymbol{\beta}}$ in the sense that

$$\sqrt{N}\big(\widehat{\boldsymbol{\beta}}_{(i)} - \boldsymbol{\beta}\big) = \sqrt{N}\big(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\big) + o_p(1) \tag{A.34}$$

implying that $\widehat{\boldsymbol{\beta}}_{(i)}$ and $\widehat{\boldsymbol{\beta}}$ share the same asymptotic distribution. Note that $\widehat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}$ whose asymptotic distribution is reported in Chapter 1 Formula (1.9). We first show the consistency of $\widehat{\boldsymbol{\beta}}_{(i)}$. Using formula (A.1), we know that

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{(i)} - \boldsymbol{\beta} &= (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \big(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\big)^{-1}\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i \\ &= (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \big(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\big)^{-1}\widetilde{\mathbf{X}}_i'(\mathbf{I}_T - \mathbf{H}_i)^{-1}\big(\widetilde{\mathbf{u}}_i - \widetilde{\mathbf{X}}_i(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\big)\end{aligned} \tag{A.35}$$

Using Triangle Inequality in Hansen (2019, p.795),

$$\begin{aligned}\|(\widehat{\boldsymbol{\beta}}_{(i)} - \boldsymbol{\beta})\| \leq{}& \big\|(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\big\| \\ &+ \left\|\left(\frac{1}{N}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\right\|\frac{1}{N}\|\widetilde{\mathbf{X}}_i\|\|(\mathbf{I}_T - \mathbf{H}_i)^{-1}\|\big(\|\widetilde{\mathbf{u}}_i\| + \|\widetilde{\mathbf{X}}_i\|\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|\big) \\ ={}& o_p(1)\end{aligned} \tag{A.36}$$

where $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|$ is $o_p(1)$; $\big(N^{-1}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\big)^{-1} = \mathbf{S}_{XX}^{-1} + o_p(1)$ by ASM.3, *WLLN* and *Slutsky's theorem*; $\widetilde{\mathbf{X}}_i$ and $\widetilde{\mathbf{u}}_i$ are $O_p(1)$ because random variables with finite moments by ASM.4 and, therefore,

$N^{-1}\|\widetilde{\mathbf{X}}_i\| = O_p(N^{-1})$; and

$$\begin{aligned}
\|(\mathbf{I}_T - \mathbf{H}_i)^{-1}\| &= \|\mathbf{I}_T + \widetilde{\mathbf{X}}_i(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} - \widetilde{\mathbf{X}}_i'\widetilde{\mathbf{X}}_i)^{-1}\widetilde{\mathbf{X}}_i'\| \\
&\leq \|\mathbf{I}_T\| + \|\widetilde{\mathbf{X}}_i\|^2\|(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} - \widetilde{\mathbf{X}}_i'\widetilde{\mathbf{X}}_i)^{-1}\| \\
&= \sqrt{T} + \frac{1}{N}\|\widetilde{\mathbf{X}}_i\|^2\left\|\left(\frac{1}{N}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} - \frac{1}{N}\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{X}}_i\right)^{-1}\right\|
\end{aligned} \tag{A.37}$$

is O(1) because the first term on the right-hand-side, $\sqrt{T}$, is O(1) without a remainder term, and the second component is bounded above by $o_p(1)$ random variable with a similar argument as in (A.12). As a result, $\widehat{\boldsymbol{\beta}}_{(i)}$ is a consistent estimator of the true value of the parameter $\boldsymbol{\beta}$ as $\widehat{\boldsymbol{\beta}}_{(i)} = \boldsymbol{\beta} + o_p(1)$ from (A.36). Therefore, removing one unit does not have an impact on the estimates of the true value of the parameter as the cross-sectional units increase to infinity.

We now show that the estimators $\widehat{\boldsymbol{\beta}}_{(i)}$ and $\widehat{\boldsymbol{\beta}}$ have the same distribution.

$$\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \sqrt{N}(\widehat{\boldsymbol{\beta}}_{(i)} - \boldsymbol{\beta}) = \left(\frac{1}{N}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\frac{1}{\sqrt{N}}\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i \tag{A.38}$$

$$= \left(\frac{1}{N}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\frac{1}{\sqrt{N}}\widetilde{\mathbf{X}}_i'(\mathbf{I}_T - \mathbf{H}_i)^{-1}\left(\widetilde{\mathbf{u}}_i - \widetilde{\mathbf{X}}_i(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right) \tag{A.39}$$

Using Reverse Triangle Inequality in Hansen (2019, p.795) and Formula (A.39),

$$\begin{aligned}
\|\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \sqrt{N}(\widehat{\boldsymbol{\beta}}_{(i)} - \boldsymbol{\beta})\| &\leq \sqrt{N}\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\| + \sqrt{N}\|\widehat{\boldsymbol{\beta}}_{(i)} - \boldsymbol{\beta}\| \\
&\leq \left\|\left(\frac{1}{N}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\right\|\frac{1}{\sqrt{N}}\left\|\widetilde{\mathbf{X}}_i\right\|\|(\mathbf{I}_T - \mathbf{H}_i)^{-1}\|\left(\|\widetilde{\mathbf{u}}_i\| + \|\widetilde{\mathbf{X}}_i\|\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|\right) \\
&= o_p(1)
\end{aligned} \tag{A.40}$$

where the first component of (A.40) is $\left(\mathbf{S}_{XX}^{-1} + o_p(1)\right)$ as $N^{-1}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} \xrightarrow{p} \mathbb{E}(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}) = \mathbf{S}_{XX} > 0$ by the Central Limit Theorem; the second term is ,$O_p(N^{1/r-1/2}) = O_p(1)$ for $r \geq 2$ under ASM.4.i; the third component is $O(1)$ by (A.37); and the last quantity in parenthesis is $O_p(1)$. Therefore,

$\left\|\sqrt{N}\big(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\big)-\sqrt{N}\big(\widehat{\boldsymbol{\beta}}_{(i)}-\boldsymbol{\beta}\big)\right\|$ is bounded above by a $o_p(1)$ random variable, and we can conclude that $\widehat{\boldsymbol{\beta}}_{(i)}$ and $\widehat{\boldsymbol{\beta}}$ share the same asymptotic distribution as

$$\sqrt{N}\big(\widehat{\boldsymbol{\beta}}_{(i)}-\boldsymbol{\beta}\big)=\sqrt{N}\big(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\big)+o_p(1) \tag{A.41}$$

Now, we find the distribution of $\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\beta}}_{(i)}$. Recall that

$$\begin{aligned}\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\beta}}_{(i)}&=\big(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\big)^{-1}\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i\\&=\big(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\big)^{-1}\widetilde{\mathbf{X}}_i'(\mathbf{I}_T-\mathbf{H}_i)^{-1}\big(\widetilde{\mathbf{u}}_i-\widetilde{\mathbf{X}}_i(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\big)\end{aligned} \tag{A.42}$$

Using Shwarz Inequality and Triangle Inequality in Hansen (2019, p.795)

$$\begin{aligned}\big\|\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\beta}}_{(i)}\big\|&=\big\|\big(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\big)^{-1}\widetilde{\mathbf{X}}_i'(\mathbf{I}_T-\mathbf{H}_i)^{-1}\big(\widetilde{\mathbf{u}}_i-\widetilde{\mathbf{X}}_i(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\big)\big\|\\&\le\big\|N\big(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\big)^{-1}\big\|N^{-1}\big\|\widetilde{\mathbf{X}}_i\big\|\big\|(\mathbf{I}_T-\mathbf{H}_i)^{-1}\big\|\big(\|\widetilde{\mathbf{u}}_i\|+\|\widetilde{\mathbf{X}}_i\|\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|\big)\\&=o_p(1)\end{aligned} \tag{A.43}$$

where $\big\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\big\|\xrightarrow{p}0$; $\big(N^{-1}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\big)^{-1}=\mathbf{S}_{XX}^{-1}+o_p(1)$ by ASM.3, *WLLN* and *Slutsky's theorem*; $\widetilde{\mathbf{X}}_i$ and $\widetilde{\mathbf{u}}_i$ are $O_p(1)$ because random variables with finite moments by ASM.4 and, therefore, $N^{-1}\big\|\widetilde{\mathbf{X}}_i\big\|=O_p(N^{-1})$ for $r\ge 2$; and $\big\|(\mathbf{I}_T-\mathbf{H}_i)^{-1}\big\|$ is bounded above by $o_p(1)$ random variable as in (A.37).

The last steps of the proof lead to the derivation of the exact distribution of $\mathrm{C}_{ii}(\widehat{\boldsymbol{\beta}})$. In Formula (A.41) we showed that $\widehat{\boldsymbol{\beta}}_{(i)}$ and $\widehat{\boldsymbol{\beta}}$ follow the same normal asymptotic distribution but we have not yet defined the asymptotic variance of $\widehat{\boldsymbol{\beta}}_{(i)}$. We derive the exact variance of $\widehat{\boldsymbol{\beta}}_{(i)}$ under model assumptions ASM.1–ASM.4. Note that

$$\begin{aligned}\mathbb{E}\Big[\widehat{\mathbf{u}}_i(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'|\widetilde{\mathbf{X}}_i\Big] &= \mathbb{E}\Big[\big(\widetilde{\mathbf{u}}_i-\widetilde{\mathbf{X}}_i(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\big)(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'|\widetilde{\mathbf{X}}_i\Big]\\ &= \mathbb{E}\Big[\widetilde{\mathbf{u}}_i(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'|\widetilde{\mathbf{X}}_i\Big]-\mathbb{E}\Big[\widetilde{\mathbf{X}}_i(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'|\widetilde{\mathbf{X}}_i\Big]\\ &= -\widetilde{\mathbf{X}}_i\mathrm{Var}\big(\widehat{\boldsymbol{\beta}}|\widetilde{\mathbf{X}}\big) \qquad \text{(A.44)}\\ \mathbb{E}\Big[(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\widehat{\mathbf{u}}_i'|\widetilde{\mathbf{X}}_i\Big] &= -\mathrm{Var}\big(\widehat{\boldsymbol{\beta}}|\widetilde{\mathbf{X}}\big)\widetilde{\mathbf{X}}_i' \qquad \text{(A.45)}\end{aligned}$$

where $\mathrm{Var}\big(\widehat{\boldsymbol{\beta}}|\widetilde{\mathbf{X}}\big)=\big(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\big)^{-1}\widetilde{\mathbf{X}}'\overline{\boldsymbol{\Sigma}}_N\widetilde{\mathbf{X}}\big(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\big)^{-1}$ as derived in Chapter 1.4 and since $\widehat{\boldsymbol{\beta}}$ and the error term are statistically independent, and

$$\begin{aligned}\mathbb{E}\Big[\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'|\widetilde{\mathbf{X}}_i\Big] &= \mathbb{E}\Big[\big(\widetilde{\mathbf{u}}_i-\widetilde{\mathbf{X}}_i(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\big)\big(\widetilde{\mathbf{u}}_i-\widetilde{\mathbf{X}}_i(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\big)'|\widetilde{\mathbf{X}}_i\Big]\\ &= \mathbb{E}\Big[\widetilde{\mathbf{u}}_i\widetilde{\mathbf{u}}_i'|\widetilde{\mathbf{X}}_i\Big]-\mathbb{E}\Big[\widetilde{\mathbf{u}}_i(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'|\widetilde{\mathbf{X}}_i\Big]\widetilde{\mathbf{X}}_i'-\widetilde{\mathbf{X}}_i\mathbb{E}\Big[(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\widetilde{\mathbf{u}}_i'|\widetilde{\mathbf{X}}_i\Big]\\ &\quad+\widetilde{\mathbf{X}}_i\mathbb{E}\Big[(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'|\widetilde{\mathbf{X}}_i\Big]\widetilde{\mathbf{X}}_i'\\ &= \boldsymbol{\Sigma}_i+\widetilde{\mathbf{X}}_i\mathrm{Var}\big(\widehat{\boldsymbol{\beta}}|\widetilde{\mathbf{X}}\big)\widetilde{\mathbf{X}}_i' \qquad \text{(A.46)}\end{aligned}$$

where $\boldsymbol{\Sigma}_i=\mathbb{E}\big(\widetilde{\mathbf{u}}_i\widetilde{\mathbf{u}}_i'|\widetilde{\mathbf{X}}_i\big)$ by ASM.2.ii. Let $\mathrm{A}(\widetilde{\mathbf{X}}_i)\equiv\big(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\big)^{-1}\widetilde{\mathbf{X}}_i'$ and $\mathrm{B}(\widetilde{\mathbf{X}}_i)=\mathrm{A}(\widetilde{\mathbf{X}}_i)\mathbf{M}_i^{-1}\widetilde{\mathbf{X}}_i$. Combining this information, the conditional finite sample variance of $\widehat{\boldsymbol{\beta}}_{(i)}$ is

$$\begin{aligned}\mathrm{Var}\big(\widehat{\boldsymbol{\beta}}_{(i)}|\widetilde{\mathbf{X}}_i\big) &= \mathbb{E}\Big\{\big(\widehat{\boldsymbol{\beta}}_{(i)}-\boldsymbol{\beta}\big)\big(\widehat{\boldsymbol{\beta}}_{(i)}-\boldsymbol{\beta}\big)'|\widetilde{\mathbf{X}}_i\Big\}\\ &= \mathbb{E}\Big\{\Big((\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})-\mathrm{A}(\widetilde{\mathbf{X}}_i)\widehat{\mathbf{v}}_i\Big)\Big((\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})-\mathrm{A}(\widetilde{\mathbf{X}}_i)\widehat{\mathbf{v}}_i\Big)'\Big|\widetilde{\mathbf{X}}_i\Big\}\\ &= \mathbb{E}\Big\{\Big((\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'-(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\widehat{\mathbf{v}}_i'\mathrm{A}(\widetilde{\mathbf{X}}_i)'-\mathrm{A}(\widetilde{\mathbf{X}}_i)\widehat{\mathbf{v}}_i(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\\ &\quad+\mathrm{A}(\widetilde{\mathbf{X}}_i)\widehat{\mathbf{v}}_i\widehat{\mathbf{v}}_i'\mathrm{A}(\widetilde{\mathbf{X}}_i)'\Big)\Big|\widetilde{\mathbf{X}}_i\Big\}\\ &= \mathbb{E}\Big\{(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'|\widetilde{\mathbf{X}}_i\Big\}-\mathbb{E}\Big\{(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\widehat{\mathbf{u}}_i'|\widetilde{\mathbf{X}}_i\Big\}\mathbf{M}_i^{-1}\mathrm{A}(\widetilde{\mathbf{X}}_i)'\\ &\quad-\mathrm{A}(\widetilde{\mathbf{X}}_i)\mathbf{M}_i^{-1}\mathbb{E}\Big\{\widehat{\mathbf{u}}_i(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'|\widetilde{\mathbf{X}}_i\Big\}+\mathrm{A}(\widetilde{\mathbf{X}}_i)\mathbf{M}_i^{-1}\mathbb{E}\Big\{\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'|\widetilde{\mathbf{X}}_i\Big\}\mathbf{M}_i^{-1}\mathrm{A}(\widetilde{\mathbf{X}}_i)'\end{aligned}$$

$$= \operatorname{Var}\big(\widehat{\boldsymbol{\beta}}|\widetilde{\mathbf{X}}\big) + \operatorname{Var}\big(\widehat{\boldsymbol{\beta}}|\widetilde{\mathbf{X}}\big)\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\mathrm{A}(\widetilde{\mathbf{X}}_i)' + \mathrm{A}(\widetilde{\mathbf{X}}_i)\mathbf{M}_i^{-1}\widetilde{\mathbf{X}}_i\operatorname{Var}\big(\widehat{\boldsymbol{\beta}}|\widetilde{\mathbf{X}}\big) \tag{A.47}$$

$$+\mathrm{A}(\widetilde{\mathbf{X}}_i)\mathbf{M}_i^{-1}\Big\{\boldsymbol{\Sigma}_i + \widetilde{\mathbf{X}}_i\operatorname{Var}\big(\widehat{\boldsymbol{\beta}}|\widetilde{\mathbf{X}}\big)\widetilde{\mathbf{X}}_i'\Big]\mathbf{M}_i^{-1}\mathrm{A}(\widetilde{\mathbf{X}}_i)' \tag{A.48}$$

$$= \operatorname{Var}\big(\widehat{\boldsymbol{\beta}}|\widetilde{\mathbf{X}}\big) + \operatorname{Var}\big(\widehat{\boldsymbol{\beta}}|\widetilde{\mathbf{X}}\big)\mathrm{B}(\widetilde{\mathbf{X}}_i)' + \mathrm{B}(\widetilde{\mathbf{X}}_i)\operatorname{Var}\big(\widehat{\boldsymbol{\beta}}|\widetilde{\mathbf{X}}\big) \tag{A.49}$$

$$+\mathrm{A}(\widetilde{\mathbf{X}}_i)\mathbf{M}_i^{-1}\boldsymbol{\Sigma}_i\mathbf{M}_i^{-1}\mathrm{A}(\widetilde{\mathbf{X}}_i)' + \mathrm{B}(\widetilde{\mathbf{X}}_i)\operatorname{Var}\big(\widehat{\boldsymbol{\beta}}|\widetilde{\mathbf{X}}\big)\mathrm{B}(\widetilde{\mathbf{X}}_i)' \tag{A.50}$$

where $\widehat{\mathbf{u}}_i = \widetilde{\mathbf{u}}_i - \widetilde{\mathbf{X}}_i\big(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\big)$ and using the unbiasedness of $\widehat{\boldsymbol{\beta}}$, the conditional exogeneity of the disturbances, and Equations (A.44)–(A.45). Under the model assumptions and using THM 6.6 in Hansen (2019, p.182), the matrices $\mathrm{A}_N(\widetilde{\mathbf{X}}_i) \equiv \big(N^{-1}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\big)^{-1}N^{-1}\widetilde{\mathbf{X}}_i' \xrightarrow{p} \mathbf{S}_{XX}^{-1}\mathbf{0}$ and $\mathrm{B}_N(\widetilde{\mathbf{X}}_i) \equiv \mathrm{A}_N(\widetilde{\mathbf{X}}_i)\mathbf{M}_i^{-1}\widetilde{\mathbf{X}}_i \xrightarrow{p} \mathbf{0}$ as $N \to \infty$ and T fixed. From Result (A.50), the asymptotic variance of $\sqrt{N}\big(\widehat{\boldsymbol{\beta}}_{(i)} - \boldsymbol{\beta}\big)$ is

$$\operatorname{Avar}\Big(\sqrt{N}\big(\widehat{\boldsymbol{\beta}}_{(i)} - \boldsymbol{\beta}\big)\Big) = N\operatorname{Var}\big(\widehat{\boldsymbol{\beta}}_{(i)}|\widetilde{\mathbf{X}}_i\big) \equiv \mathbf{V}_{\widehat{\boldsymbol{\beta}}} \tag{A.51}$$

where $\mathbf{V}_{\widehat{\boldsymbol{\beta}}} = \mathbf{S}_{XX}^{-1}\boldsymbol{\Sigma}\mathbf{S}_{XX}^{-1}$, as shown in Chapter 1.4. Therefore, we can conclude that $\widehat{\boldsymbol{\beta}}_{(i)}$ has identical limiting distribution as $\widehat{\boldsymbol{\beta}}$. The distance between $\widehat{\boldsymbol{\beta}}_{(i)}$ and $\widehat{\boldsymbol{\beta}}$ vanishes as panel units grow and, hence, the influence exerted by unit i is null.

We follow the same reasoning as above to derive the distribution of $\widehat{\boldsymbol{\beta}}_{(i,j)}$. Adding and subtracting $\boldsymbol{\beta}$ from Formula (A.33),

$$\widehat{\boldsymbol{\beta}}_{(i,j)} - \boldsymbol{\beta} = \big(\widehat{\boldsymbol{\beta}}_{(i)} - \boldsymbol{\beta}\big) - \big(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\big)^{-1}\big(\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\mathbf{H}_{ij} + \widetilde{\mathbf{X}}_j'\big)$$
$$\big(\mathbf{M}_j - \mathbf{H}_{ij}'\mathbf{M}_i^{-1}\mathbf{H}_{ij}\big)^{-1}\big(\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i + \widehat{\mathbf{u}}_j\big) \tag{A.52}$$

where $\widehat{\mathbf{u}}_l = \big(\widetilde{\mathbf{u}}_l - \widetilde{\mathbf{X}}_l(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\big)$ with $l \in \{i, j\}$, and $\mathbf{H}_{ij} = \widetilde{\mathbf{X}}_i(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}_j'$. Using Triangle Inequality in Hansen (2019, p.795),

$$\|\widehat{\boldsymbol{\beta}}_{(i,j)} - \boldsymbol{\beta}\| \leq \|\widehat{\boldsymbol{\beta}}_{(i)} - \boldsymbol{\beta}\| \tag{A.53}$$

$$+ \left\| \left(\frac{1}{N} \widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} \right)^{-1} \right\| \tag{A.54}$$

$$\left(\frac{1}{N} \|\widetilde{\mathbf{X}}_i\| \|\mathbf{M}_i^{-1}\| \|\mathbf{H}_{ij}\| + \frac{1}{N} \|\widetilde{\mathbf{X}}_j\| \right) \tag{A.55}$$

$$\left\| \left(\mathbf{M}_j - \mathbf{H}_{ij}'\mathbf{M}_i^{-1}\mathbf{H}_{ij} \right)^{-1} \right\| \tag{A.56}$$

$$\left(\|\mathbf{H}_{ij}\| \|\mathbf{M}_i^{-1}\| \|\widehat{\mathbf{u}}_i\| + \|\widehat{\mathbf{u}}_j\| \right) \tag{A.57}$$

where the term on the right-hand-side of (A.53) is $o_p(1)$ by result (A.36); (A.54) is $\mathbf{S}_{XX}^{-1} + o_p(1)$ as $N \to \infty$ and T fixed; (A.55) is $O_p(1)$ because of $N^{-1}\|\widetilde{\mathbf{X}}_l\| = O_p(N^{-1})$ for $l \in \{i, j\}$, $\|\mathbf{M}_i^{-1}\| = O(1)$ by (A.37), and $\|\mathbf{H}_{ij}\| = N^{-1/2}\|\widetilde{\mathbf{X}}_i\| \|(N^{-1}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\| N^{-1/2}\|\widetilde{\mathbf{X}}_j\| = o_p(1)$; component (A.56) is $O_p(1)$ because $\left(\mathbf{M}_j - \mathbf{H}_{ij}'\mathbf{M}_i^{-1}\mathbf{H}_{ij} \right)^{-1} = ß_T + o_p(1)$ since $\mathbf{M}_j = \mathbf{I}_T - N^{-1/2}\widetilde{\mathbf{X}}_j(N^{-1}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}N^{-1/2}\widetilde{\mathbf{X}}_j' = \mathbf{I}_T + o_p(1)$; and, (A.57) is $O_p(1)$ because $\|\widehat{\mathbf{u}}_l\| = \|\widetilde{\mathbf{u}}_l\| + \|\widetilde{\mathbf{X}}_l\| \|(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\| = O_p(1)$ by ASM.4. Therefore, $\|\widehat{\boldsymbol{\beta}}_{(i,j)} - \boldsymbol{\beta}\|$ would be $o_p(1)$ from (A.53) and (A.54) and, hence, $\widehat{\boldsymbol{\beta}}_{(i,j)} = \boldsymbol{\beta} + o_p(1)$.

We verify that the estimators $\widehat{\boldsymbol{\beta}}_{(i,j)}$ and $\widehat{\boldsymbol{\beta}}$ have the same asymptotic distribution. Adding and subtracting $\boldsymbol{\beta}$ from the first line of Formula (A.33),

$$\begin{aligned}
&\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \sqrt{N}(\widehat{\boldsymbol{\beta}}_{(i,j)} - \boldsymbol{\beta}) \\
&\quad = \left(\frac{1}{N} \widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} \right)^{-1} \frac{1}{\sqrt{N}} \widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i \\
&\quad\quad + \left(\frac{1}{N} \widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} \right)^{-1} \frac{1}{\sqrt{N}} (\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\mathbf{H}_{ij} + \widetilde{\mathbf{X}}_j') \\
&\quad\quad \left(\mathbf{M}_j - \mathbf{H}_{ij}'\mathbf{M}_i^{-1}\mathbf{H}_{ij} \right)^{-1} \left(\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i + \widehat{\mathbf{u}}_j \right)
\end{aligned} \tag{A.58}$$

Using Reverse Triangle Inequality in Hansen (2019, p.795) and Formula (A.58),

$$
\begin{aligned}
&\left\|\sqrt{N}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})-\sqrt{N}(\widehat{\boldsymbol{\beta}}_{(i,j)}-\boldsymbol{\beta})\right\| \\
&\quad \leq \left\|\left(\frac{1}{N}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\right\|\left\|\frac{1}{\sqrt{N}}\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i\right\| && \text{(A.59)}\\
&\quad + \left\|\left(\frac{1}{N}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\right\|\left\|\frac{1}{\sqrt{N}}\left(\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\mathbf{H}_{ij}+\widetilde{\mathbf{X}}_j'\right)\right\| && \text{(A.60)}\\
&\qquad \left\|\left(\mathbf{M}_j-\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\mathbf{H}_{ij}\right)^{-1}\right\| && \text{(A.61)}\\
&\qquad \left\|\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i+\widehat{\mathbf{u}}_j\right\| && \text{(A.62)}\\
&\quad = o_p(1)
\end{aligned}
$$

where (A.59) is $o_p(1)$ by result (A.40); the overall quantity in (A.60) is $o_p(1)$ because the first component is $\mathbf{S}_{XX}^{-1}+o_p(1)$ and multiplies a quantity that is $O_p(1)$ as in (A.54), noting that $N^{-1/2}\|\widetilde{\mathbf{X}}_i\| = O_p(N^{1/r-1/2})$ and, hence, $O_p(1)$ with $r \geq 2$; (A.61) is $O_p(1)$ as in (A.36); (A.62) is $O_p(1)$ as in (A.57). Therefore, $\widehat{\boldsymbol{\beta}}_{(i,j)}$ and $\widehat{\boldsymbol{\beta}}$ share the same asymptotic distribution

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}_{(i,j)}-\boldsymbol{\beta}) = \sqrt{N}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})+o_p(1) \tag{A.63}$$

Last step consists in deriving the joint distribution of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}_{(i,j)}$. Rearranging Equation (A.33), we obtain

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\beta}}_{(i,j)} &= \left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i \\
&\quad + \left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\left(\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\mathbf{H}_{ij}+\widetilde{\mathbf{X}}_j'\right) \\
&\qquad \left(\mathbf{M}_j-\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\mathbf{H}_{ij}\right)^{-1}\left(\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i+\widehat{\mathbf{u}}_j\right) && \text{(A.64)}
\end{aligned}
$$

Using Shwarz Inequality and Triangle Inequality in Hansen (2019, p.795)

$$\|\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\beta}}_{(i,j)}\| \leq \left\| \left(\frac{1}{N}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1} \right\| \left\| \frac{1}{N}\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i \right\| \tag{A.65}$$

$$+\left\| \left(\frac{1}{N}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1} \right\| \left\| \frac{1}{N}\left(\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\mathbf{H}_{ij}+\widetilde{\mathbf{X}}_j'\right) \right\| \tag{A.66}$$

$$\left\| \left(\mathbf{M}_j-\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\mathbf{H}_{ij}\right)^{-1} \right\| \tag{A.67}$$

$$\left\| \mathbf{H}_{ij}'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i+\widehat{\mathbf{u}}_j \right\| \tag{A.68}$$

where the left-hand-side is $o_p(1)$ following the reasoning used in (A.53)-(A.57). Thus, the difference $\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\beta}}_{(i,j)} \xrightarrow{p} \mathbf{0}$ as $N \to \infty$ and T fixed.

Noting that $\widehat{\boldsymbol{\beta}}_{(i,j)}-\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}_{(i)}-\boldsymbol{\beta}-\mathrm{D}_1(.)\mathrm{D}_2(.)$ with $\mathrm{D}_1\left(\widetilde{\mathbf{X}}_l\right)=\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\left(\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\mathbf{H}_{ij}+\widetilde{\mathbf{X}}_j'\right)\left(\mathbf{M}_j-\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\mathbf{H}_{ij}\right)^{-1}$, $\mathrm{D}_2\left(\widetilde{\mathbf{X}}_l,\widetilde{\mathbf{u}}_l,\boldsymbol{\beta}\right)=\left(\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i+\widehat{\mathbf{u}}_j\right)$, the exact finite sample variance of $\widehat{\boldsymbol{\beta}}_{(i,j)}$ under the true model assumptions ASM.1–ASM.4 is

$$\mathrm{Var}\left(\widehat{\boldsymbol{\beta}}_{(i,j)}|\widetilde{\mathbf{X}}_i,\widetilde{\mathbf{X}}_j\right)=\mathbb{E}\left\{\left(\widehat{\boldsymbol{\beta}}_{(i,j)}-\boldsymbol{\beta}\right)\left(\widehat{\boldsymbol{\beta}}_{(i,j)}-\boldsymbol{\beta}\right)'|\widetilde{\mathbf{X}}_i,\widetilde{\mathbf{X}}_j\right\}$$

$$=\mathbb{E}\left\{\left(\widehat{\boldsymbol{\beta}}_{(i)}-\boldsymbol{\beta}\right)\left(\widehat{\boldsymbol{\beta}}_{(i)}-\boldsymbol{\beta}\right)'|\widetilde{\mathbf{X}}_i,\widetilde{\mathbf{X}}_j\right\} \tag{A.69}$$

$$-\mathbb{E}\left\{\left(\widehat{\boldsymbol{\beta}}_{(i)}-\boldsymbol{\beta}\right)\mathrm{D}_2(.)'\mathrm{D}_1(.)'\middle|\widetilde{\mathbf{X}}_i,\widetilde{\mathbf{X}}_j\right\} \tag{A.70}$$

$$-\mathbb{E}\left\{\mathrm{D}_1(.)\mathrm{D}_2(.)\left(\widehat{\boldsymbol{\beta}}_{(i)}-\boldsymbol{\beta}\right)'\middle|\widetilde{\mathbf{X}}_i,\widetilde{\mathbf{X}}_j\right\} \tag{A.71}$$

$$+\mathbb{E}\left\{\mathrm{D}_1(.)\mathrm{D}_2(.)\mathrm{D}_2(.)'\mathrm{D}_1(.)'\middle|\widetilde{\mathbf{X}}_i,\widetilde{\mathbf{X}}_j\right\} \tag{A.72}$$

where $\mathrm{D}_1(.)\equiv\mathrm{D}_1\left(\widetilde{\mathbf{X}}_l\right)$, $\mathrm{D}_2(.)\equiv\mathrm{D}_2\left(\widetilde{\mathbf{X}}_l,\widetilde{\mathbf{u}}_l,\boldsymbol{\beta}\right)$, and $\widehat{\mathbf{u}}_i=\widetilde{\mathbf{u}}_i-\widetilde{\mathbf{X}}_i\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)$. Component (A.69) is equal (A.49)-(A.50). Component (A.70) is the transpose of (A.71), which is

$$
\begin{aligned}
&\mathbb{E}\Big[\mathrm{D}_1(.)\mathrm{D}_2(.)\big(\widehat{\boldsymbol{\beta}}_{(i)}-\boldsymbol{\beta}\big)'\Big|\widetilde{\mathbf{X}}_i,\widetilde{\mathbf{X}}_j\Big] \\
&\quad=\mathbb{E}\Big[\mathrm{D}_1(.)\mathrm{D}_2(.)\Big(\big(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\big)-\mathrm{A}(\widetilde{\mathbf{X}}_i)\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i\Big)'\Big|\widetilde{\mathbf{X}}_i,\widetilde{\mathbf{X}}_j\Big] && \text{(A.73)}\\
&\quad=\mathbb{E}\Big[\mathrm{D}_1(.)\mathrm{D}_2(.)\big(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\big)'\Big|\widetilde{\mathbf{X}}_i,\widetilde{\mathbf{X}}_j\Big]-\mathbb{E}\Big[\mathrm{D}_1(.)\mathrm{D}_2(.)\widehat{\mathbf{u}}_i'\mathbf{M}_i^{-1}\mathrm{A}(\widetilde{\mathbf{X}}_i)'\Big|\widetilde{\mathbf{X}}_i,\widetilde{\mathbf{X}}_j\Big] && \text{(A.74)}\\
&\quad=\mathrm{D}_1(.)\Big\{\mathbb{E}\Big[\big(\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i+\widehat{\mathbf{u}}_j\big)\big(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\big)'\Big|\widetilde{\mathbf{X}}_i,\widetilde{\mathbf{X}}_j\Big] && \text{(A.75)}\\
&\quad-\mathbb{E}\Big[\big(\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i+\widehat{\mathbf{u}}_j\big)\widehat{\mathbf{u}}_i'\Big|\widetilde{\mathbf{X}}_i,\widetilde{\mathbf{X}}_j\Big]\mathbf{M}_i^{-1}\mathrm{A}(\widetilde{\mathbf{X}}_i)'\Big\} && \text{(A.76)}\\
&\quad=\mathrm{D}_1(.)\Big\{\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\mathbb{E}\Big[\widehat{\mathbf{u}}_i\big(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\big)'\Big|\widetilde{\mathbf{X}}_i\Big]+\mathbb{E}\Big[\widehat{\mathbf{u}}_j\big(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\big)'\Big|\widetilde{\mathbf{X}}_j\Big] && \text{(A.77)}\\
&\quad-\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\mathbb{E}\Big[\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\Big|\widetilde{\mathbf{X}}_i\Big]\mathbf{M}_i^{-1}\mathrm{A}(\widetilde{\mathbf{X}}_i)'-\mathbb{E}\Big[\widehat{\mathbf{u}}_j\widehat{\mathbf{u}}_i'\Big|\widetilde{\mathbf{X}}_i,\widetilde{\mathbf{X}}_j\Big]\mathbf{M}_i^{-1}\mathrm{A}(\widetilde{\mathbf{X}}_i)'\Big\} && \text{(A.78)}
\end{aligned}
$$

where (A.73) is derived by plugging (A.1) and $\mathrm{A}(\widetilde{\mathbf{X}}_i)\equiv\big(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\big)^{-1}\widetilde{\mathbf{X}}_i'$; in (A.77), $\mathbb{E}\Big[\widehat{\mathbf{u}}_l\big(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\big)'\Big|\widetilde{\mathbf{X}}_l\Big]=-\widetilde{\mathbf{X}}_l\mathrm{Var}\big(\widehat{\boldsymbol{\beta}}|\widetilde{\mathbf{X}}\big)$ with $l\in\{i,j\}$ by (A.44); in (A.78), $\mathbb{E}\Big[\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\Big|\widetilde{\mathbf{X}}_i\Big]=\boldsymbol{\Sigma}_i+\widetilde{\mathbf{X}}_i\mathrm{Var}\big(\widehat{\boldsymbol{\beta}}|\widetilde{\mathbf{X}}\big)\widetilde{\mathbf{X}}_i'$ by (A.46), and $\mathbb{E}\Big[\widehat{\mathbf{u}}_j\widehat{\mathbf{u}}_i'\Big|\widetilde{\mathbf{X}}_i,\widetilde{\mathbf{X}}_j\Big]=\widetilde{\mathbf{X}}_j\mathrm{Var}\big(\widehat{\boldsymbol{\beta}}|\widetilde{\mathbf{X}}\big)\widetilde{\mathbf{X}}_i'$ because the error term is independent across individuals by assumption and uncorrelated with the estimated parameter of interest. Focusing on component (A.72),

$$
\begin{aligned}
&\mathbb{E}\Big[\mathrm{D}_1(.)\mathrm{D}_2(.)\mathrm{D}_2(.)'\mathrm{D}_1(.)'\Big|\widetilde{\mathbf{X}}_i,\widetilde{\mathbf{X}}_j\Big]\\
&\quad=\mathrm{D}_1(.)\Big\{\mathbb{E}\Big[\big(\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i+\widehat{\mathbf{u}}_j\big)\big(\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i+\widehat{\mathbf{u}}_j\big)'\Big|\widetilde{\mathbf{X}}_i,\widetilde{\mathbf{X}}_j\Big]\Big\}\mathrm{D}_1(.)' && \text{(A.79)}\\
&\quad=\mathrm{D}_1(.)\Big\{\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\mathbb{E}\Big[\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\Big|\widetilde{\mathbf{X}}_i\Big]\mathbf{M}_i^{-1}\mathbf{H}_{ij}'+\mathbb{E}\Big[\widehat{\mathbf{u}}_j\widehat{\mathbf{u}}_j'\Big|\widetilde{\mathbf{X}}_j\Big] && \text{(A.80)}\\
&\quad+2\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\mathbb{E}\Big[\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_j'\Big|\widetilde{\mathbf{X}}_i\Big]\Big\}\mathrm{D}_1(.)' && \text{(A.81)}
\end{aligned}
$$

where the covariance between $\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i$ and $\widehat{\mathbf{u}}_j$ is equal to the quanity in (A.81) by the independence across individuals (ASM.1) and the mean of residuals being zero. Also, $\mathbb{E}\Big[\widehat{\mathbf{u}}_l\widehat{\mathbf{u}}_l'\Big|\widetilde{\mathbf{X}}_l\Big]=\boldsymbol{\Sigma}_l+\widetilde{\mathbf{X}}_l\mathrm{Var}\big(\widehat{\boldsymbol{\beta}}|\widetilde{\mathbf{X}}\big)\widetilde{\mathbf{X}}_l'$ by (A.46), and $\mathbb{E}\Big[\widehat{\mathbf{u}}_j\widehat{\mathbf{u}}_i'\Big|\widetilde{\mathbf{X}}_i,\widetilde{\mathbf{X}}_j\Big]=\widetilde{\mathbf{X}}_j\mathrm{Var}\big(\widehat{\boldsymbol{\beta}}|\widetilde{\mathbf{X}}\big)\widetilde{\mathbf{X}}_i'$ because the error term is independent across individuals by ASM.1 and uncorrelated with the estimated parameter of interest.

Under the model assumptions, $N^{-1}\widetilde{\mathbf{X}}_l' \xrightarrow{p} \mathbf{0}$ wich drives $\mathrm{D}_1(.)_N \xrightarrow{p} \mathbf{0}$ and, therefore, (A.69)–(A.72) are asymptotically zero. Then, the variance of $\sqrt{N}\big(\widehat{\boldsymbol{\beta}}_{(i)}-\boldsymbol{\beta}\big)$ converges in probability to $\mathbf{V}_{\widehat{\boldsymbol{\beta}}}$ due to the only component in (A.69) that does not vanishes towards zero. The L2O estimator $\widehat{\boldsymbol{\beta}}_{(i,j)}$ has the same limiting distribution as the *within-group* estimator, $\widehat{\boldsymbol{\beta}}$.

A.7.2 Conditional Influence

Consider and rearranging Equation (2.43) as follows

$$\begin{aligned}\sqrt{N}\big(\widehat{\boldsymbol{\beta}}_{(i,j)}-\boldsymbol{\beta}\big)-\sqrt{N}\big(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\big)= \\ -\left(\frac{1}{N}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\Bigg\{\frac{1}{\sqrt{N}}\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i \\ +\frac{1}{\sqrt{N}}\big(\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\mathbf{H}_{ij}+\widetilde{\mathbf{X}}_j'\big)\big(\mathbf{M}_j-\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\mathbf{H}_{ij}\big)^{-1}\big(\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i+\widehat{\mathbf{u}}_j\big)\Bigg\}\end{aligned} \tag{A.82}$$

and Equation (A.38) with respect to j

$$\sqrt{N}\big(\widehat{\boldsymbol{\beta}}_{(j)}-\boldsymbol{\beta}\big)-\sqrt{N}\big(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\big)=-\left(\frac{1}{N}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\frac{1}{\sqrt{N}}\widetilde{\mathbf{X}}_j'\mathbf{M}_j^{-1}\widehat{\mathbf{u}}_j \tag{A.83}$$

such that the difference of (A.82) and (A.38) is as follows

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}_{(i,j)}-\boldsymbol{\beta})-\sqrt{N}(\widehat{\boldsymbol{\beta}}_{(j)}-\boldsymbol{\beta})=\left(\frac{1}{N}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1} \tag{A.84}$$

$$\Bigg\{\frac{1}{\sqrt{N}}\widetilde{\mathbf{X}}_j'\mathbf{M}_j^{-1}\widehat{\mathbf{u}}_j-\frac{1}{\sqrt{N}}\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i \tag{A.85}$$

$$-\frac{1}{\sqrt{N}}\big(\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\mathbf{H}_{ij}+\widetilde{\mathbf{X}}_j'\big)\big(\mathbf{M}_j-\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\mathbf{H}_{ij}\big)^{-1}\big(\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i+\widehat{\mathbf{u}}_j\big)\Bigg\} \tag{A.86}$$

Then,

$$\begin{aligned}
&\left\|\sqrt{N}(\widehat{\boldsymbol{\beta}}_{(i,j)}-\boldsymbol{\beta})-\sqrt{N}(\widehat{\boldsymbol{\beta}}_{(j)}-\boldsymbol{\beta})\right\| \\
&\quad\leq\left\|\left(\frac{1}{N}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\right\|\left\{\left\|\frac{1}{\sqrt{N}}\widetilde{\mathbf{X}}_j'\mathbf{M}_j^{-1}\widehat{\mathbf{u}}_j\right\|+\left\|\frac{1}{\sqrt{N}}\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i\right\|\right. && \text{(A.87)}\\
&\quad\left.+\left\|\frac{1}{\sqrt{N}}(\widetilde{\mathbf{X}}_i'\mathbf{M}_i^{-1}\mathbf{H}_{ij}+\widetilde{\mathbf{X}}_j')\right\|\left\|\left(\mathbf{M}_j-\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\mathbf{H}_{ij}\right)^{-1}\right\|\left\|(\mathbf{H}_{ij}'\mathbf{M}_i^{-1}\widehat{\mathbf{u}}_i+\widehat{\mathbf{u}}_j)\right\|\right\} && \text{(A.88)}
\end{aligned}$$

is $o_p(1)$ by Results (A.54)-(A.57) and (A.40). Therefore,

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{(i,j)}-\boldsymbol{\beta}\right)=\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{(j)}-\boldsymbol{\beta}\right)+o_p(1) \qquad \text{(A.89)}$$

which is asymptotically equivalent to $\widehat{\boldsymbol{\beta}}$ by Result (A.41) relatively to unit j. Noting that $\sqrt{N}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\xrightarrow{d}\mathcal{N}\left(\mathbf{0},\mathbf{S}_{XX}^{-1}\boldsymbol{\Sigma}\mathbf{S}_{XX}^{-1}\right)$.

A.8 Kitagawa-Blinder-Oaxaca Decomposition

The Blinder-Oaxaca decomposition of a generic outcome variable into three components (*endowment effect*, *coefficients effect*, and *interaction effect*) is derived as follows. Consider two groups (e.g., men and women) and the group-specific linear model

$$y_g = \widetilde{\mathbf{X}}_g' \boldsymbol{\beta}_g + \varepsilon_g \qquad g \in \{M, W\} \tag{A.90}$$

where y_g is the logarithm of wages; $\widetilde{\mathbf{X}}_g$ is a $k \times 1$ vector containing the predictors and a constant term (including female dominance, low segregation, and their interaction, working region, socio-demographic factors, occupation controls); $\boldsymbol{\beta}_g$ is a vector of slope parameters and the intercept; ε_g is the nuisance term. The difference in the mean outcome of the two groups can be expressed as follows

$$\mathbb{E}(y_M) - \mathbb{E}(y_W) = \mathbb{E}(\widetilde{\mathbf{X}}_M)' \boldsymbol{\beta}_M - \mathbb{E}(\widetilde{\mathbf{X}}_W)' \boldsymbol{\beta}_W \tag{A.91}$$

because $\mathbb{E}(\varepsilon_g) = 0$ and $\mathbb{E}(\boldsymbol{\beta}_g) = \boldsymbol{\beta}_g$, under classical linear model assumptions. The decomposition from the viewpoint of women (W) is as follows

$$\begin{aligned}
&\mathbb{E}(y_M) - \mathbb{E}(y_W) \\
&= \mathbb{E}(\widetilde{\mathbf{X}}_M)' \boldsymbol{\beta}_M - \mathbb{E}(\widetilde{\mathbf{X}}_W)' \boldsymbol{\beta}_W \\
&= [\mathbb{E}(\widetilde{\mathbf{X}}_M) - \mathbb{E}(\widetilde{\mathbf{X}}_W)]' \boldsymbol{\beta}_W + \mathbb{E}(\widetilde{\mathbf{X}}_M)'(\boldsymbol{\beta}_M - \boldsymbol{\beta}_W) \\
&= \underbrace{[\mathbb{E}(\widetilde{\mathbf{X}}_M) - \mathbb{E}(\widetilde{\mathbf{X}}_W)]' \boldsymbol{\beta}_W}_{\text{endowment effect}} + \underbrace{\mathbb{E}(\widetilde{\mathbf{X}}_W)'(\boldsymbol{\beta}_M - \boldsymbol{\beta}_W)}_{\text{coefficients effect}} + \underbrace{[\mathbb{E}(\widetilde{\mathbf{X}}_M) - \mathbb{E}(\widetilde{\mathbf{X}}_W)]'(\boldsymbol{\beta}_M - \boldsymbol{\beta}_W)}_{\text{interaction effect}}
\end{aligned} \tag{A.92}$$

The *endowment effect* measures the expected change in women's mean outcome if they had men's predictor levels. The *coefficients effect* component measures the expected change in women's mean outcome if women had men's coefficients. The *interaction effect* captures those differences in endowments and coefficients that exist simultaneously between the two group.

Appendix B

Tables

Table B.1. *Single hypothesis test, heteroskedasticity*

	Heteroskedasticity ($\gamma = 2$)											
	PB	RP	RMSE	PB	RP	RMSE	PB	RP	RMSE	PB	RP	RMSE
(N,T)		(25, 2)			(50, 2)			(150, 2)			(500, 2)	
PHC0	0.713	0.516	0.305	0.625	0.407	0.263	0.450	0.223	0.147	0.316	0.098	0.081
PHC3	-0.083	0.017	0.035	-0.138	0.023	0.058	-0.014	0.029	0.005	0.084	0.027	0.021
PHC6	-0.042	0.020	0.018	-0.123	0.025	0.052	-0.010	0.030	0.003	0.085	0.027	0.022
PHCjk	-0.039	0.018	0.017	-0.119	0.024	0.050	-0.010	0.030	0.003	0.085	0.027	0.022
		(25, 5)			(50, 5)			(150, 5)			(500, 5)	
PHC0	0.578	0.337	0.224	0.473	0.238	0.160	0.338	0.112	0.089	0.225	0.062	0.042
PHC3	-0.121	0.022	0.047	-0.024	0.027	0.008	0.066	0.026	0.017	0.099	0.033	0.018
PHC6	-0.098	0.023	0.038	-0.024	0.027	0.008	0.069	0.026	0.018	0.100	0.033	0.018
PHCjk	-0.086	0.023	0.033	-0.010	0.029	0.003	0.069	0.026	0.018	0.100	0.033	0.018
		(25, 10)			(50, 10)			(150, 10)			(500, 10)	
PHC0	0.472	0.219	0.024	0.395	0.154	0.117	0.277	0.079	0.062	0.184	0.052	0.027
PHC3	-0.033	0.022	0.011	0.052	0.027	0.015	0.098	0.029	0.022	0.105	0.035	0.015
PHC6	-0.012	0.025	0.004	0.061	0.029	0.018	0.101	0.030	0.023	0.106	0.035	0.015
PHCjk	-0.006	0.024	0.002	0.063	0.029	0.019	0.101	0.030	0.023	0.106	0.035	0.015
		(25, 20)			(50, 20)			(150, 20)			(500, 20)	
PHC0	0.385	0.138	0.112	0.308	0.088	0.075	0.211	0.056	0.037	0.131	0.052	0.014
PHC3	0.029	0.021	0.008	0.076	0.024	0.019	0.096	0.031	0.017	0.084	0.042	0.009
PHC6	0.049	0.026	0.014	0.085	0.027	0.021	0.099	0.031	0.017	0.086	0.042	0.009
PHCjk	0.052	0.025	0.015	0.086	0.026	0.021	0.099	0.031	0.017	0.085	0.042	0.009

The number of replications is 10,000. The random variable associated with slope parameter β_1 is contaminated with leverage points and drives heteroskedasticity. PB: Proportional Bias. Positive values indicate by how much the standard error underestimates the "true" standard error. RP: Rejection Probability of 5%-level t-test on β_1 (i.e., size of test). RMSE: Root Mean Squared Error.

Table B.2. *Single hypothesis test, homoskedasticity*

	Homoskedasticity ($\gamma = 0$)											
	PB	RP	RMSE	PB	RP	RMSE	PB	RP	RMSE	PB	RP	RMSE
(N,T)		(25, 2)			(50, 2)			(150, 2)			(500, 2)	
PHC0	0.369	0.204	0.043	0.348	0.192	0.014	0.174	0.116	0.002	0.058	0.067	0.000
PHC3	-0.411	0.028	0.048	-0.374	0.040	0.015	-0.140	0.049	0.002	-0.049	0.046	0.000
PHC6	-0.328	0.036	0.038	-0.346	0.044	0.014	-0.131	0.050	0.002	-0.046	0.046	0.000
PHCjk	-0.361	0.030	0.042	-0.353	0.041	0.014	-0.135	0.049	0.002	-0.048	0.046	0.000
		(25, 5)			(50, 5)			(150, 5)			(500, 5)	
PHC0	0.324	0.160	0.008	0.205	0.118	0.002	0.074	0.075	0.000	0.017	0.055	0.000
PHC3	-0.310	0.040	0.007	-0.160	0.050	0.002	-0.062	0.047	0.000	-0.027	0.046	0.000
PHC6	-0.284	0.045	0.007	-0.148	0.052	0.002	-0.059	0.048	0.000	-0.026	0.047	0.000
PHCjk	-0.273	0.043	0.006	-0.146	0.052	0.002	-0.059	0.048	0.000	-0.026	0.047	0.000
		(25, 10)			(50, 10)			(150, 10)			(500, 10)	
PHC0	0.197	0.113	0.002	0.125	0.095	0.001	0.041	0.066	0.000	0.025	0.057	0.000
PHC3	-0.179	0.043	0.002	-0.071	0.050	0.000	-0.032	0.050	0.000	0.002	0.051	0.000
PHC6	-0.156	0.048	0.002	-0.060	0.053	0.000	-0.028	0.051	0.000	0.003	0.051	0.000
PHCjk	-0.151	0.047	0.002	-0.059	0.052	0.000	-0.028	0.051	0.000	0.003	0.051	0.000
		(25, 20)			(50, 20)			(150, 20)			(500, 20)	
PHC0	0.114	0.084	0.001	0.065	0.068	0.001	0.019	0.053	0.000	0.008	0.052	0.000
PHC3	-0.097	0.047	0.001	-0.043	0.048	0.001	-0.020	0.045	0.000	-0.004	0.049	0.000
PHC6	-0.076	0.053	0.001	-0.033	0.050	0.000	-0.016	0.046	0.000	-0.003	0.049	0.000
PHCjk	-0.074	0.050	0.000	-0.033	0.050	0.000	-0.016	0.046	0.000	-0.003	0.049	0.000

The number of replications is 10,000. The random variable associated with slope parameter β_1 is contaminated with leverage points and drives heteroskedasticity. PB: Proportional Bias. Positive values indicate by how much the standard error underestimates the "true" standard error. RP: Rejection Probability of 5%-level t-test on β_1 (i.e., size of test). RMSE: Root Mean Squared Error.

Table B.3. *Joint hypothesis test, heteroskedasticity*

	Heteroskedasticity ($\gamma = 2$)							
	Wald Stats	RP	Wald Stats	RP	Wald Stats	RP	Wald Stats	RP
(N,T)	(25,2)		(50, 2)		(150, 2)		(500, 2)	
PHC0	3341260.250	0.923	29613.070	0.820	6.124	0.510	1.775	0.210
PHC3	117.439	0.112	9.203	0.150	1.513	0.129	1.105	0.076
PHC6	66.965	0.264	9.947	0.281	2.495	0.223	1.606	0.189
PHCjk	124.333	0.115	9.505	0.154	1.525	0.130	1.107	0.077
	(25, 5)		(50, 5)		(150, 5)		(500, 5)	
PHC0	1293.329	0.768	19.074	0.569	2.138	0.263	1.365	0.115
PHC3	9.351	0.169	1.964	0.140	1.175	0.093	1.081	0.060
PHC6	10.219	0.303	2.872	0.248	1.737	0.197	1.709	0.217
PHCjk	10.377	0.178	2.019	0.145	1.184	0.094	1.084	0.060
	(25, 10)		(50, 10)		(150, 10)		(500, 10)	
PHC0	16.202	0.568	3.348	0.355	1.540	0.153	1.210	0.076
PHC3	2.157	0.152	1.353	0.110	1.086	0.065	1.053	0.044
PHC6	3.565	0.256	2.107	0.213	1.639	0.191	1.864	0.264
PHCjk	2.317	0.163	1.389	0.113	1.094	0.067	1.055	0.045
	(25, 20)		(50, 20)		(150, 20)		(500, 20)	
PHC0	4.169	0.375	1.972	0.213	1.328	0.098	1.145	0.064
PHC3	1.568	0.119	1.201	0.082	1.077	0.052	1.055	0.046
PHC6	2.438	0.234	1.820	0.202	1.773	0.230	2.311	0.390
PHCjk	1.658	0.130	1.228	0.087	1.085	0.054	1.057	0.046

The number of replications is 10,000. Tested hypothesis $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 1$. Random variables associated with slope parameters β_1 and β_3 are contaminated with leverage points. All random variables drive heteroskedasticity. RP: Rejection Probability of 5%-level t-test (i.e., size of test).

Table B.4. *Joint hypothesis test, homoskedasticity*

	Homoskedasticity ($\gamma = 0$)							
	Wald Stats	RP	Wald Stats	RP	Wald Stats	RP	Wald Stats	RP
(N,T)	(25,2)		(50, 2)		(150, 2)		(500, 2)	
PHC0	118.278	0.724	12.974	0.575	2.298	0.298	1.341	0.136
PHC3	1.444	0.103	1.478	0.132	1.235	0.112	1.098	0.082
PHC6	24.939	0.323	10.828	0.427	5.985	0.471	4.638	0.495
PHCjk	1.521	0.109	1.514	0.136	1.244	0.113	1.100	0.083
	(25, 5)		(50, 5)		(150, 5)		(500, 5)	
PHC0	8.066	0.522	2.845	0.340	1.463	0.162	1.144	0.087
PHC3	1.855	0.155	1.399	0.132	1.134	0.086	1.050	0.066
PHC6	10.577	0.470	7.094	0.484	5.191	0.491	5.164	0.661
PHCjk	1.957	0.166	1.431	0.136	1.142	0.088	1.053	0.066
	(25, 10)		(50, 10)		(150, 10)		(500, 10)	
PHC0	3.252	0.343	1.853	0.216	1.265	0.110	1.075	0.069
PHC3	1.588	0.139	1.286	0.110	1.102	0.078	1.026	0.058
PHC6	8.413	0.492	5.917	0.476	4.834	0.545	6.700	0.845
PHCjk	1.665	0.147	1.313	0.115	1.109	0.079	1.028	0.059
	(25, 20)		(50, 20)		(150, 20)		(500, 20)	
PHC0	2.104	0.227	1.465	0.143	1.152	0.079	1.055	0.066
PHC3	1.437	0.121	1.196	0.088	1.067	0.060	1.029	0.059
PHC6	6.842	0.234	5.376	0.508	5.584	0.705	10.257	0.986
PHCjk	1.501	0.489	1.221	0.094	1.075	0.062	1.031	0.060

The number of replications is 10,000. Tested hypothesis $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 1$. Random variables associated with slope parameters β_1 and β_3 are contaminated with leverage points. All random variables drive heteroskedasticity. RP: Rejection Probability of 5%-level t-test (i.e., size of test).

Table B.5. *List of replicated articles*

Article	Cit.	Table	Specification	N	NT	Change Significance
Acemoglu et al. (2008)	1,834	2	(2), (7)	150; 127	945; 457	N, N
Schularick et al. (2012)	2,852	2	(2), (3)	14	1,272	Y, Y
Égert (2016)	84	3	(1), (1), (1)	34	844	Y, N, Y
Berka et al. (2018)	58	4	(2a), (2b), (2c)	9	117	N, N, Y

Note: The listed papers are published in AER. The number of citations comes from Google Scholar in date 17 May 2021.

Table B.6. *Replicates of Acemoglu et al.'s (2008) Table 2*

	Base sample, 1960–2000	
	Five-year data (2)	Ten-year data (7)
Dep. variable: Democracy		
Democracy_{t-1}	0.379	-0.025
	(0.0467)***	(0.0747)
	[0.0478]***	[0.0774]
	{0.0479}***	{0.0778}
	⟨0.0479⟩***	⟨0.0778⟩
lnGDP_{t-1}	0.010	0.053
	(0.0317)	(0.0564)
	[0.0324]	[0.0576]
	{0.0325}	{0.0579}
	⟨0.0325⟩	⟨0.0579⟩
Observations	945	457
Groups	150	127
Time Periods	8	4
Test for groupwise heteroskedasticity		
$\chi^2(N)$	5.3e+34	1.5e+11
p-value	0.0000	0.0000

*This table replicates Acemoglu et al.'s (2008) Table 2. The dependent variable is "freedom house" used as a measure of democracy. The coefficients are estimated using OLS in a model with fixed effects. The number of time series is the maximum number of periods used in the estimation procedure once the first t periods have been removed from the sample due to identification and multicollinearity. Cluster-robust standard errors (Arellano's (1987)) are in parenthesis; PHC6 standard errors in brackets; PHC3 in curly brackets; jackknife standard errors in angle brackets. Original standard errors (not reported here) use the asymptotic uncorrected formula. The test for groupwise heteroskedasticity is a modified Wald test for fixed effect regression model which tests the null hypothesis of constant variance of the disturbances against the alternative of non-constant variance. Significance levels: *** 0.01 ** 0.05 * 0.10.*

Table B.7. *Replicates of Schularick and Taylor's (2012) Table 3*

	(2)	(3)
Dep. variable: Financial Crisis (binary)		
Δlog(loans/P)$_{t-1}$	-0.0273	-0.0489
	(0.1285)	(0.1051)
	[0.1360]	[0.1133]
	{0.1358}	{0.1170}
	⟨0.1358⟩	⟨0.1170⟩
Δlog(loans/P)$_{t-2}$	0.302	0.302
	(0.1153)**	(0.0860)***
	[0.1235]**	[0.0936]***
	{0.1232}**	{0.0950}***
	⟨0.1232⟩**	⟨0.0950⟩***
Δlog(loans/P)$_{t-3}$	0.0478	0.00134
	(0.1295)	(0.0886)
	[0.1390]	[0.0913]
	{0.1390}	{0.1018}
	⟨0.1390⟩	⟨0.1018⟩
Δlog(loans/P)$_{t-4}$	0.00213	0.0346
	(0.0664)	(0.0822)
	[0.0717]	[0.0839]
	{0.0714}	{0.0923}
	⟨0.0714⟩	⟨0.0923⟩
Δlog(loans/P)$_{t-5}$	0.0928	0.136
	(0.0355)**	(0.0317)***
	[0.0358]**	[0.0321]***
	{0.0357}**	{0.0341}***
	⟨0.0357⟩**	⟨0340⟩***
Year FE	No	Yes
Observations	1,272	1,272
Groups	14	14
Test for groupwise heteroskedasticity		
$\chi^2(N)$	23.08	135.32
p-value	0.0590	0.0000

*This table replicates Specifications (2) and (3) of Table 3 in Schularick and Taylor (2012). The coefficients are estimated using OLS in a model with fixed effects. Cluster-robust standard errors (Arellano's (1987)) are in parenthesis; PHC6 standard errors in brackets; PHC3 in curly brackets; jackknife standard errors in angle brackets. The test for groupwise heteroskedasticity is a modified Wald test for fixed effect regression model which tests the null hypothesis of constant variance of the disturbances against the alternative of non-constant variance. Significance levels: *** 0.01 ** 0.05 * 0.10.*

Table B.8. *Replicates of Égert's (2016) Table R1*

	(1)	(2)	(3)
Dep. variable: Multi-Factor Productivity (MFP)			
ETCR overall	-0.058		
	(0.022)**		
	[0.0252]**		
	{0.0254}**		
	⟨0.0254⟩**		
ETCR entry barriers		-0.039	
		(0.0144)**	
		[0.0161]**	
		{0.0162}**	
		⟨0.0162⟩**	
ETCR public ownership			-0.050
			(0.0225)**
			[0.0273]*
			{0.0273}*
			⟨0.0273⟩*
Openness size adjusted	0.006	0.006	0.007
	(0.0028)**	(0.0027)**	(0.0025)***
	[0.0032]*	[0.0031]*	[0.0028]**
	{0.0032}*	{0.0031}*	{0.0028}**
	⟨0.0032⟩*	⟨0.0031⟩*	⟨0.0028⟩**
Business exp. on R&D	0.048	0.043	0.072
	(0.043)	(0.0444)	(0.0387)*
	[0.0499]	[0.0500]	[0.0451]
	{0.0500}	{0.0509}	{0.0451}
	⟨0.0500⟩	⟨0.0509⟩	⟨0.0451⟩
Human capital	0.167	0.320	0.380
	(0.338)	(0.3203)	(0.3084)
	[0.3918]	[0.3665]	[0.3522]
	{0.3930}	{0.3684}	{0.3526}
	⟨0.3930⟩	⟨0.3683⟩	⟨0.3525⟩
Output gap	0.010	0.010	0.009
	(0.0014)***	(0.0016)***	(0.0012)***
	[0.0016]***	[0.0018]***	[0.0013]***
	{0.0016}***	{0.0018}***	{0.0013}***
	⟨0.0016⟩***	⟨0.0018⟩***	⟨0.0013⟩***
Time FE	No	No	No
Observations	844	844	844
Groups	34	34	34
Test for groupwise heteroskedasticity			
$\chi^2(N)$	2037.36	2693.33	1605.04
p-value	0.0000	0.0000	0.0000

*This table replicates baseline regression results in Table R1 in Égert's (2016) Online Appendix. Cluster-robust standard errors (Arellano's (1987)) are in parenthesis; PHC6 standard errors in brackets; PHC3 in curly brackets; jackknife standard errors in angle brackets. The test for groupwise heteroskedasticity is a modified Wald test for fixed effect regression model which tests the null hypothesis of constant variance of the disturbances against the alternative of non-constant variance. Significance levels: *** 0.01 ** 0.05 * 0.10.*

Table B.9. *Replicates of Berka et al.'s (2018) Table 4*

	Fixed effects		
	(2a)	(2b)	(2c)
Dep. variable: log real exchange rate			
TFP	-0.10		
	(0.1582)		
	[0.1688]		
	{0.1691}		
	⟨0.1690⟩		
TFP_T		0.003	0.18
		(0.1254)	(0.1418)
		[0.1297]	[0.1418]
		{0.1357}	{0.2087}
		⟨0.1357⟩	⟨0.2069⟩
TFP_N		-0.36	-0.36
		(0.2255)	(0.2637)
		[0.2331]	[0.2637]
		{0.2377}	{0.3431}
		⟨0.2377⟩	⟨0.3405⟩
RULC_T			0.46
			(0.1077)***
			[0.1077]***
			{0.2133}*
			⟨0.2113⟩*
Observations	117	117	117
Groups	9	9	9
Test for groupwise heteroskedasticity			
$\chi^2(N)$	165.20	230.40	10.95
p-value	0.0000	0.0000	0.2792

*This table replicates Table 4 in Berka et al. (2018). The coefficients are estimated using OLS in a model with fixed effects. Cluster-robust standard errors (Arellano's (1987)) are in parenthesis; PHC6 standard errors in brackets; PHC3 in curly brackets; jackknife standard errors in angle brackets. Original standard errors, omitted from the table, are calculated using period weighting (PCSE) and degree-of-freedom correction. The test for groupwise heteroskedasticity is a modified Wald test for fixed effect regression model which tests the null hypothesis of constant variance of the disturbances against the alternative of non-constant variance. Significance levels: *** 0.01 ** 0.05 * 0.10.*

Table B.10. *Summary statistics, female sample*

Variable	Count	Mean	Standard Deviation	Minimum	Maximum
Demographic characteristics					
Women	1,788,945				
Natives	1,788,945	0.85	0.36	0	1
EEA	1,788,945	0.05	0.22	0	1
non-EEA	1,788,945	0.10	0.30	0	1
Age	1,788,945	39.85	13.52	16	64
Black	1,788,945	0.03	0.16	0	1
Asian	1,788,945	0.05	0.22	0	1
Other ethnicity	1,788,945	0.04	0.19	0	1
Muslim	1,788,945	0.04	0.19	0	1
Christian	1,788,945	0.56	0.50	0	1
Other religions	1,788,945	0.16	0.37	0	1
Socio-economic factors					
In couple	1,788,945	0.51	0.50	0	1
With dependent children	1,788,945	0.37	0.48	0	1
Years of Education	1,768,266	13.21	3.05	9	20
Experience	1,637,066	23.74	13.27	0	59
Training	798,713	0.33	0.47	0	1
In labour force	1,788,945	0.70	0.46	0	1
Employed	1,246,572	0.94	0.24	0	1
Part-time work	1,462,879	0.43	0.50	0	1
Public sector	1,166,118	0.33	0.47	0	1
Permanent job	45,755	0.73	0.44	0	1
Hours	1,152,745	30.89	13.36	0	196
Weekly hours	1,152,372	3.32	0.52	0	5.28
Remote work	1,788,945	0.03	0.17	0	1
Benefit	1,788,945	0.46	0.50	0	1
Female dominance	1,161,370	0.69	0.46	0	1
Low segregation	1,161,370	0.27	0.45	0	1

Notes: If applicable, the number of hours includes usual hours of paid overtime to total usual hours worked in main job.

Table B.11. *Summary statistics, male sample*

Variable	Count	Mean	Standard Deviation	Minimum	Maximum
Demographic characteristics					
Men	1,544,280				
Natives	1,544,280	0.87	0.34	0	1
EEA	1,544,280	0.04	0.21	0	1
non-EEA	1,544,280	0.09	0.28	0	1
Age	1,544,280	40.14	14.26	16	64
Black	1,544,280	0.02	0.15	0	1
Asian	1,544,280	0.05	0.21	0	1
Other ethnicity	1,544,280	0.03	0.18	0	1
Muslim	1,544,280	0.04	0.19	0	1
Christian	1,544,280	0.52	0.50	0	1
Other religions	1,544,280	0.16	0.37	0	1
Socio-economic factors					
In couple	1,544,280	0.50	0.50	0	1
With dependent children	1,544,280	0.28	0.45	0	1
Years of Education	1,520,860	13.11	2.87	9	20
Experience	1,398,836	24.30	13.86	0	59
Training	829,306	0.32	0.47	0	1
In labour force	1,544,280	0.79	0.41	0	1
Employed	1,218,593	0.92	0.27	0	1
Part-time work	1,325,507	0.12	0.32	0	1
Public sector	1,118,726	0.13	0.34	0	1
Permanent job	43,059	0.73	0.44	0	1
Hours	1,100,480	40.33	13.58	0	196
Remote work	1,544,280	0.05	0.23	0	1
Benefit	1,544,280	0.20	0.40	0	1
Female dominance	1,113,792	0.35	0.48	0	1
Low segregation	1,113,792	0.26	0.44	0	1

Notes: If applicable, the number of hours includes usual hours of paid overtime to total usual hours worked in main job.

Table B.12. *Reason for part-time work, female sample*

	Percentage (%)
Reason for part-time work	
Student or at school	11.72
Ill or disabled	2.16
Could not find full-time job	10.43
Did not want full-time job	75.69
Total	100
Among those who did not want full-time job	
Looking after children	70.79
Looking after incapacitated adult	3.81
Some other reason	25.41
Total	100

Table B.13. *List of female- and male-dominated sectors*

Female dominated sectors	Male dominated sectors
G - Distribution	A - Agriculture, forestry & fishing
I - Accommodation & food services	B - Mining & quarrying
L - Real estate services	C - Manufacturing
P - Education	D - Electricity, gas & air con supply
Q - Health & social work	E - Water supply, sewerage & waste
S - Other service activities	F - Construction
T - Households as employers	H - Transport & storage
	J - Information & communication
	K - Financial & insurance services
	M - Professional, scientific & technical activities
	N - Admin & support services
	R - Arts, entertainment & recreation

Notes: Sectors labelled as O - Public admin & defense and U - Extra territorial are removed from the sample because their contracts and wages highly differ from other sectors for the nature of the job.

Table B.14. *List of high and low segregated sectors*

High segregated sectors	Low segregated sectors
C - Manufacturing	A - Agriculture, forestry & fishing
F - Construction	B - Mining & quarrying
H - Transport & Storage	D - Electricity, gas & air con supply
I - Accommodation & food services	E - Water supply, sewerage & waste
J - Information & communication	G - Distribution
M - Professional, scientific & technical activities	K - Financial & insurance services
P - Education,	L - Real estate services
Q - Health & social work	N - Admin & support services
S - Other service activities	R - Arts, entertainment & recreation
	T - Households as employers

Notes: Sectors labelled as O - Public admin & defense and U - Extra territorial are removed from the sample because their contracts and wages highly differ from other sectors for the nature of the job.

Table B.15. *Probit for in the labour force, marginal effects*

	Male sample		Female sample	
	2005-2020	2020	2005-2020	2020
	(1)	(2)	(3)	(4)
Dep. var: In the Labour Force				
EEA	0.026***	0.046***	-0.017***	-0.000
	(0.001)	(0.007)	(0.002)	(0.007)
non-EEA	-0.007***	0.007	-0.097***	-0.075***
	(0.001)	(0.007)	(0.002)	(0.007)
In couple	0.038***	0.017***	-0.034***	-0.043***
	(0.001)	(0.003)	(0.001)	(0.003)
With dep. children	0.065***	0.070***	0.023***	0.043***
	(0.001)	(0.003)	(0.001)	(0.005)
Middle Education	-0.005***	-0.001	0.012***	0.000
	(0.001)	(0.003)	(0.001)	(0.004)
High Education	-0.017***	-0.006	-0.023***	-0.017***
	(0.001)	(0.004)	(0.001)	(0.005)
Benefit	-0.239***	-0.253***	-0.256***	-0.239***
	(0.001)	(0.003)	(0.001)	(0.004)
Time FE	Yes	Yes	Yes	Yes
Region Controls	Yes	Yes	Yes	Yes
Socio-demographic Controls	Yes	Yes	Yes	Yes
Observations	1,381,712	6,1380	1,622,063	73,930

*Notes: Data from UK Labour Force Survey (LFS). All models are estimated using a Probit for binary dependent variables. Marginal effects are reported with their significance levels. Robust standard errors in parenthesis. Significance levels: p<0.01 ***, p<0.05 **, p<0.1 *.*

Table B.16. *Contribution of individual components of KBO decomposition, pooled sample*

Year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Endowments															
ln(hours)	-0.121***	-0.104***	-0.159***	-0.129***	-0.131***	-0.130***	-0.110***	-0.071***	-0.110***	-0.162***	-0.126***	-0.113***	-0.129***	-0.087***	-0.071***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.014)	(0.014)	(0.013)	(0.013)	(0.014)	(0.014)	(0.014)	(0.014)	(0.015)	(0.016)	(0.015)
ln(hours)^2	0.120***	0.098***	0.161***	0.127***	0.135***	0.131***	0.108***	0.072***	0.107***	0.167***	0.130***	0.115***	0.135***	0.082***	0.065***
	(0.014)	(0.014)	(0.015)	(0.014)	(0.015)	(0.015)	(0.015)	(0.015)	(0.016)	(0.016)	(0.015)	(0.015)	(0.017)	(0.017)	(0.017)
EEA	-0.000	-0.000	0.000	0.000	0.000	0.000*	0.000	0.000*	0.000	0.000	-0.000	0.000	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
non-EEA	-0.000*	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age	0.011	0.028***	0.029**	0.004	0.004	0.002	0.009*	0.001	0.003	0.007	0.005	0.000	0.004	0.003	0.015*
	(0.007)	(0.008)	(0.010)	(0.010)	(0.003)	(0.002)	(0.004)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)	(0.006)
Age^2	0.014	-0.009	0.002	0.024*	0.004	0.006	-0.003	0.006	0.001	-0.001	-0.001	0.002	-0.002	-0.001	-0.005
	(0.008)	(0.009)	(0.011)	(0.011)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
experience	0.011*	0.011*	0.010	0.025***	0.005*	0.003	0.004	0.008**	0.004	0.004	0.005*	0.006*	0.003	0.005*	-0.002
	(0.005)	(0.005)	(0.006)	(0.007)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)
Experience^2	-0.038***	-0.035***	-0.040***	-0.053***	-0.011**	-0.007	-0.008**	-0.011**	-0.006*	-0.008*	-0.006*	-0.005	-0.003	-0.004	-0.006*
	(0.005)	(0.006)	(0.007)	(0.007)	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)	(0.002)	(0.003)	(0.003)
Midle educ	-0.001	-0.001	-0.001	-0.001	-0.001*	0.000	-0.001	-0.002*	-0.000	-0.001	-0.001*	-0.000	-0.000	-0.001*	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
High educ	0.003***	0.001*	0.002***	0.001	0.002**	0.002**	0.001	0.001	0.001	0.001	-0.000	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
years educ	0.010**	0.002	-0.003	-0.010*	0.001	0.004	-0.001	-0.012*	-0.006	-0.012**	-0.006	-0.017***	-0.020***	-0.020***	-0.024***
	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
years educ^2	-0.004	0.001	0.006	0.012**	0.002	0.001	0.004	0.012**	0.007	0.012**	0.006	0.016***	0.019***	0.018***	0.021***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)
Training	-0.001***	-0.002***	-0.002***	-0.001**	-0.001***	-0.001***	-0.001**	-0.001***	-0.000*	-0.001***	-0.000	-0.000	-0.000	-0.000	-0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
In couple	-0.000	0.000	0.000	0.000	0.000	0.001**	0.000	0.000	0.001	0.000	0.000	0.001	0.001*	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
With dep. children	0.004***	0.004***	0.003***	0.004***	0.002***	0.002***	0.001**	0.002***	0.002***	0.002**	0.001**	0.001*	0.002**	0.002**	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)
In couple with dep. children	0.000	-0.000	0.000	-0.000	0.000	0.000	0.000	0.001	0.001	0.001*	0.001*	0.001	0.001	0.001*	0.002*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)
fml dominated sector	0.051***	0.049***	0.049***	0.050***	0.047***	0.053***	0.055***	0.056***	0.054***	0.060***	0.049***	0.052***	0.054***	0.056***	0.056***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
low segregation	0.000	0.003***	0.002**	0.001	0.001	0.001*	0.000	0.002**	0.001*	0.000	0.000	-0.000	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
fml dom sector and low segr.	-0.000	-0.003***	-0.001	-0.001	-0.001	-0.001*	-0.001	-0.002***	-0.002**	-0.001	-0.000	-0.000	-0.001*	-0.001*	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Public	-0.014***	-0.013***	-0.013***	-0.018***	-0.017***	-0.016***	-0.018***	-0.016***	-0.015***	-0.013***	-0.010***	-0.008***	-0.006***	-0.007***	-0.013***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Coefficients															
ln(hours)	1.912***	1.232***	2.177***	2.654***	0.864***	1.598***	1.459***	1.235***	0.764**	2.611***	1.836***	1.858***	2.106***	2.204***	1.274***
	(0.241)	(0.229)	(0.233)	(0.245)	(0.229)	(0.222)	(0.241)	(0.245)	(0.241)	(0.273)	(0.271)	(0.280)	(0.314)	(0.307)	(0.350)
ln(hours)^2	-1.087***	-0.789***	-1.274***	-1.470***	-0.637***	-0.971***	-0.884***	-0.796***	-0.530***	-1.447***	-1.072***	-1.078***	-1.246***	-1.251***	-0.837***
	(0.123)	(0.120)	(0.123)	(0.130)	(0.122)	(0.118)	(0.128)	(0.129)	(0.128)	(0.143)	(0.142)	(0.146)	(0.163)	(0.162)	(0.188)
EEA	-0.001	0.000	0.000	-0.001	-0.001	-0.000	0.000	-0.000	-0.003*	-0.003*	-0.004*	-0.001	-0.002	-0.002	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
non-EEA	0.000	0.000	-0.000	-0.001	-0.001	-0.004**	-0.003	-0.001	-0.003	-0.002	0.000	-0.001	-0.003	-0.003	0.001
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Age	-0.154	0.137	-0.035	0.605	0.435	0.494	-0.029	0.944**	0.254	-0.103	0.159	0.960**	-0.295	0.247	0.020
	(0.305)	(0.289)	(0.304)	(0.317)	(0.345)	(0.331)	(0.333)	(0.340)	(0.340)	(0.355)	(0.339)	(0.351)	(0.352)	(0.363)	(0.432)
Age^2	-0.175	-0.078	0.012	-0.262	-0.043	-0.388*	0.182	-0.434**	-0.085	0.107	0.115	-0.275	0.227	0.150	-0.039
	(0.146)	(0.141)	(0.145)	(0.153)	(0.170)	(0.163)	(0.167)	(0.168)	(0.171)	(0.181)	(0.175)	(0.178)	(0.183)	(0.188)	(0.225)
Experience	0.319**	0.088	0.130	-0.076	-0.026	0.141	0.093	-0.122	0.065	0.133	-0.082	-0.295*	0.136	-0.123	0.183
	(0.117)	(0.108)	(0.116)	(0.120)	(0.134)	(0.127)	(0.125)	(0.125)	(0.125)	(0.128)	(0.117)	(0.123)	(0.120)	(0.122)	(0.150)
Experience^2	0.011	-0.001	-0.047	0.049	-0.069	0.038	-0.126*	0.070	-0.038	-0.067	-0.044	0.058	-0.107	-0.065	-0.073
	(0.049)	(0.046)	(0.048)	(0.050)	(0.058)	(0.055)	(0.056)	(0.056)	(0.057)	(0.060)	(0.057)	(0.059)	(0.060)	(0.062)	(0.075)
Midle educ	0.014**	0.004	0.006	-0.001	-0.001	0.017**	-0.005	-0.006	0.006	0.006	-0.006	-0.003	0.000	-0.009	0.001
	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)
High educ	0.015**	0.001	-0.000	-0.006	-0.007	0.006	-0.017*	-0.010	-0.001	0.001	-0.013	-0.022*	-0.000	-0.028**	-0.010
	(0.005)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)	(0.012)
years educ	0.316	0.341	0.155	0.463*	0.376	0.475*	0.548*	0.362	0.783**	0.865**	0.859**	0.786**	0.324	0.899**	0.720*
	(0.198)	(0.202)	(0.213)	(0.223)	(0.244)	(0.239)	(0.252)	(0.256)	(0.260)	(0.278)	(0.282)	(0.283)	(0.295)	(0.310)	(0.353)
years educ^2	-0.165	-0.168	-0.076	-0.206*	-0.183	-0.207	-0.243*	-0.155	-0.347**	-0.373**	-0.378**	-0.335*	-0.108	-0.400**	-0.300
	(0.090)	(0.092)	(0.097)	(0.103)	(0.113)	(0.110)	(0.117)	(0.120)	(0.122)	(0.131)	(0.133)	(0.135)	(0.140)	(0.148)	(0.170)
Training	0.010	0.003	0.011*	0.019**	0.008	-0.001	0.006*	0.004	0.007**	0.002	0.006**	0.006**	0.004	0.006**	0.003
	(0.005)	(0.005)	(0.005)	(0.006)	(0.005)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
In couple	0.042***	0.025***	0.044***	0.036***	0.018*	0.016*	0.051***	0.029***	0.040***	0.048***	0.042***	0.035***	0.033***	0.025**	0.041***
	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)
With dep. children	0.037***	0.043***	0.048***	0.043***	0.042***	0.034***	0.048***	0.036***	0.043***	0.031***	0.036***	0.031***	0.036***	0.038***	0.027**
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
In couple with dep. children	-0.004	0.001	-0.006	-0.000	-0.005	-0.001	-0.023***	-0.001	-0.012	-0.012	-0.009	-0.006	-0.006	-0.013	-0.006
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)
Works in London	0.003	0.004**	0.003*	0.002	0.002	0.003*	0.001	0.000	0.004*	-0.001	-0.000	-0.001	-0.001	-0.002	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
fml dominated sector	-0.019*	-0.026**	-0.027**	-0.014	-0.048***	-0.029**	-0.023*	-0.025**	-0.024*	-0.023*	-0.025*	-0.031**	-0.029**	-0.023*	-0.029*
	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.012)
Low segregation	-0.007	0.013**	0.004	0.002	0.008	0.014*	0.004	0.010	0.009	0.006	0.002	-0.001	0.018**	0.015*	0.002
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
fml dom sec and low segr.	0.014**	-0.001	0.004	0.002	0.008	-0.001	-0.002	-0.004	0.005	0.008	0.004	0.010	-0.003	-0.007	-0.001
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Public	-0.012**	-0.010**	-0.014***	-0.016***	-0.013**	-0.009*	-0.017***	-0.016***	-0.015**	-0.005	-0.011*	-0.004	0.001	-0.001	-0.013*
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
Observations	32,066	33,053	31,683	29,746	26,656	27,577	26,540	26,531	27,039	25,873	24,915	25,819	24,236	23,151	21,378

*Notes: Contribution of main socio-demographic characteristics, human capital attributes and sectoral indicators. Significance levels: pvalue<0.01 ***, pvalue<0.05 **, pvalue<0.1 *.*

Table B.17. *Contribution of individual components of KBO decomposition, female dominated sectors*

Year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Endowments															
ln(hours)	-0.078***	-0.044**	-0.095***	-0.075***	-0.072***	-0.038**	-0.051***	-0.048***	-0.054***	-0.046**	-0.067***	-0.040**	-0.058***	-0.048**	-0.044**
	(0.014)	(0.014)	(0.015)	(0.013)	(0.013)	(0.014)	(0.014)	(0.014)	(0.014)	(0.015)	(0.014)	(0.014)	(0.016)	(0.017)	(0.016)
$\ln(hours)^2$	0.073***	0.036*	0.095***	0.075***	0.074***	0.033*	0.046**	0.046**	0.050***	0.044**	0.065***	0.038*	0.056**	0.043*	0.034*
	(0.016)	(0.015)	(0.017)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.017)	(0.015)	(0.015)	(0.017)	(0.019)	(0.017)
EEA	-0.000	-0.000	-0.000	0.000	0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000	0.000	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
non-EEA	-0.002*	-0.002*	-0.001	-0.001	-0.001*	-0.002*	-0.001	-0.001	-0.000	-0.002**	-0.002*	-0.000	0.000	-0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Age	-0.001	0.001	0.006	-0.009	-0.004	-0.004	-0.014*	0.002	-0.001	-0.001	0.000	0.001	-0.012	0.000	-0.021*
	(0.002)	(0.001)	(0.005)	(0.006)	(0.006)	(0.005)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.007)	(0.006)	(0.009)
Age^2	0.005	0.004	0.008	0.032***	-0.004	-0.009	0.002	-0.014*	-0.011	-0.006	-0.010	-0.008	0.000	-0.007	0.006
	(0.004)	(0.004)	(0.008)	(0.009)	(0.004)	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)	(0.006)	(0.006)
Experience	-0.001	-0.002	0.002	0.010*	-0.009	-0.004	-0.005	-0.012*	-0.013*	-0.018**	-0.018***	-0.017**	-0.011*	-0.016**	0.003
	(0.002)	(0.002)	(0.002)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.006)	(0.004)
$Experience^2$	-0.005	-0.013**	-0.019***	-0.035***	0.010*	0.018**	0.012**	0.020***	0.020***	0.019**	0.024***	0.020***	0.017**	0.019***	0.010*
	(0.005)	(0.004)	(0.005)	(0.007)	(0.004)	(0.005)	(0.004)	(0.006)	(0.005)	(0.006)	(0.006)	(0.006)	(0.005)	(0.006)	(0.005)
Middle educ	0.000	-0.000	-0.000	-0.001	-0.000	0.001	-0.000	-0.000	0.000	-0.001	0.000	-0.000	0.000	-0.000	0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
High educ	0.004***	0.006***	0.007***	0.005***	0.005***	0.004**	0.004**	0.003*	0.003**	0.005**	0.002*	0.003*	0.001	0.003	0.001
	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Years educ	0.004	-0.006	-0.012	-0.013*	-0.004	-0.003	-0.015*	-0.014*	-0.016*	-0.017**	-0.013*	-0.029***	-0.026**	-0.022**	-0.031***
	(0.004)	(0.005)	(0.006)	(0.006)	(0.005)	(0.007)	(0.007)	(0.007)	(0.006)	(0.005)	(0.006)	(0.008)	(0.008)	(0.007)	(0.009)
Years educ2	0.001	0.009*	0.015**	0.014**	0.006	0.007	0.016**	0.015**	0.016**	0.016**	0.013*	0.028***	0.025***	0.020**	0.029***
	(0.003)	(0.004)	(0.005)	(0.006)	(0.004)	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.006)	(0.007)	(0.007)	(0.006)	(0.008)
Training	-0.002***	-0.003***	-0.003***	-0.002***	-0.001**	-0.001**	-0.001*	-0.001**	-0.000	-0.001*	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
In couple	0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
With dep. children	0.004**	0.007***	0.005**	0.005***	0.004**	0.003***	0.003**	0.004***	0.002*	0.003***	0.003**	0.002**	0.002**	0.002**	0.001
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
In couple with dep. children	-0.000	0.000	-0.000	-0.000	-0.000	0.000	-0.000	0.000	-0.000	0.000	0.000	0.000	0.000	0.001	0.001
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Low segr.	0.002	0.001	0.001	0.006**	0.004	0.003	0.008***	0.009***	0.005*	0.006**	0.002	0.007**	0.007**	0.011***	0.020***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Public	-0.018***	-0.018***	-0.018***	-0.026***	-0.021***	-0.020***	-0.021***	-0.022***	-0.019***	-0.017***	-0.014***	-0.013***	-0.010***	-0.012***	-0.017***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Coefficients															
ln(hours)	0.963**	-0.121	1.120***	1.969***	-0.066	1.544***	1.098***	1.373***	-0.155	0.818*	1.402***	1.484***	1.403***	1.332***	1.367**
	(0.329)	(0.342)	(0.323)	(0.341)	(0.282)	(0.313)	(0.308)	(0.368)	(0.295)	(0.332)	(0.408)	(0.395)	(0.411)	(0.404)	(0.476)
$\ln(hours)^2$	-0.546**	-0.035	-0.626***	-1.104***	-0.041	-0.758***	-0.602***	-0.690***	0.062	-0.400*	-0.713***	-0.760***	-0.697**	-0.727***	-0.682**
	(0.171)	(0.180)	(0.173)	(0.183)	(0.155)	(0.167)	(0.166)	(0.193)	(0.159)	(0.177)	(0.215)	(0.206)	(0.215)	(0.216)	(0.257)
EEA	0.001	0.000	0.000	-0.000	-0.002	0.000	-0.001	-0.002	-0.004*	-0.005**	-0.001	0.001	0.000	-0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
non-EEA	-0.000	0.001	-0.003	-0.002	-0.002	-0.005*	-0.005*	-0.000	-0.005	-0.001	0.001	-0.005	-0.006*	-0.006	0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Age	-0.018	0.749	0.486	0.586	0.321	0.561	-0.063	1.178*	0.530	-0.379	-0.128	0.551	-0.781	0.227	-0.285
	(0.419)	(0.408)	(0.419)	(0.419)	(0.472)	(0.442)	(0.429)	(0.458)	(0.440)	(0.471)	(0.435)	(0.464)	(0.464)	(0.500)	(0.587)
Age^2	-0.388	-0.340	-0.130	-0.074	0.213	-0.522*	0.336	-0.590**	-0.194	0.399	0.185	0.054	0.428	0.200	0.210
	(0.199)	(0.199)	(0.197)	(0.200)	(0.230)	(0.218)	(0.220)	(0.227)	(0.225)	(0.241)	(0.228)	(0.236)	(0.245)	(0.260)	(0.304)
Experience	0.458**	-0.058	-0.007	-0.137	0.008	0.193	0.065	-0.125	-0.012	0.160	0.020	-0.260	0.253	-0.101	0.150
	(0.165)	(0.155)	(0.162)	(0.162)	(0.186)	(0.171)	(0.162)	(0.170)	(0.162)	(0.174)	(0.150)	(0.165)	(0.158)	(0.168)	(0.210)
$Experience^2$	0.012	0.039	-0.067	-0.039	-0.234**	0.050	-0.205**	0.078	-0.039	-0.190*	-0.073	-0.033	-0.148	-0.117	-0.129
	(0.069)	(0.067)	(0.067)	(0.068)	(0.081)	(0.076)	(0.076)	(0.078)	(0.078)	(0.082)	(0.077)	(0.080)	(0.082)	(0.087)	(0.105)
Middle educ	0.018**	0.008	0.002	-0.004	-0.003	0.018**	-0.014*	-0.003	0.001	0.004	-0.001	0.000	0.007	-0.010	-0.003
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.007)	(0.008)	(0.008)	(0.009)	(0.010)
High educ	0.022**	-0.004	-0.013	-0.012	-0.019	0.006	-0.022*	-0.007	-0.005	-0.006	-0.002	-0.013	0.013	-0.022	-0.010
	(0.007)	(0.008)	(0.008)	(0.009)	(0.010)	(0.009)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)	(0.012)	(0.012)	(0.013)	(0.016)
Years educ	0.651*	0.497	0.159	0.432	0.311	0.088	0.362	-0.066	0.501	0.400	0.731	0.606	0.058	0.786	-0.031
	(0.300)	(0.314)	(0.322)	(0.327)	(0.363)	(0.346)	(0.364)	(0.367)	(0.365)	(0.395)	(0.395)	(0.403)	(0.410)	(0.449)	(0.496)
Years educ2	-0.333*	-0.265	-0.115	-0.247	-0.178	-0.072	-0.164	0.035	-0.243	-0.201	-0.363	-0.289	-0.024	-0.397	0.048
	(0.138)	(0.145)	(0.149)	(0.153)	(0.170)	(0.162)	(0.171)	(0.174)	(0.173)	(0.188)	(0.188)	(0.194)	(0.197)	(0.216)	(0.241)
Training	0.021*	0.002	0.011	0.030***	0.005	-0.002	0.005	-0.000	0.005	-0.002	0.006*	0.004	0.002	0.006	0.005
	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
In couple	0.044***	0.031**	0.051***	0.020	0.014	-0.000	0.038**	0.035**	0.043***	0.056***	0.058***	0.037**	0.015	0.022	0.044**
	(0.012)	(0.012)	(0.012)	(0.012)	(0.013)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.013)	(0.014)
With dep. children	0.031*	0.052***	0.050***	0.028*	0.045***	0.042***	0.047***	0.015	0.040***	0.027*	0.037**	0.045***	0.039**	0.033*	0.015
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.012)	(0.012)	(0.013)	(0.012)	(0.012)	(0.012)	(0.013)	(0.013)	(0.014)	(0.014)
In couple with dep. children	0.003	-0.000	-0.010	0.011	-0.009	-0.004	-0.013	0.016	-0.003	-0.008	-0.015	-0.016	-0.006	-0.004	0.007
	(0.010)	(0.011)	(0.010)	(0.011)	(0.011)	(0.010)	(0.010)	(0.010)	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)
Low segr.	0.022***	0.016**	0.015*	0.010	0.026***	0.018**	0.005	0.010	0.022***	0.027***	0.014**	0.019***	0.019***	0.009	0.006
	(0.006)	(0.006)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)
Public	-0.001	-0.002	0.008	-0.006	0.007	0.023**	-0.003	0.006	0.006	0.027***	0.010	0.010	0.019*	0.013	0.004
	(0.007)	(0.008)	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.010)
Observations	15,020	15,308	14,911	14,172	13,246	13,737	13,336	13,313	13,433	12,879	12,468	12,945	12,151	11,563	10,383

*Notes: Contribution of main socio-demographic characteristics, human capital attributes and sectoral indicators. Significance levels: pvalue<0.01 ***, pvalue<0.05 **, pvalue<0.1 *.*

Table B.18. *Contribution of individual components of KBO decomposition, male dominated sectors*

Year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Endowments															
ln(hours)	-0.164***	-0.189***	-0.213***	-0.195***	-0.221***	-0.250***	-0.183***	-0.110***	-0.216***	-0.271***	-0.209***	-0.261***	-0.216***	-0.118***	-0.111***
	(0.020)	(0.020)	(0.019)	(0.022)	(0.025)	(0.023)	(0.021)	(0.024)	(0.027)	(0.022)	(0.025)	(0.027)	(0.025)	(0.023)	(0.027)
ln(hours)^2	0.171***	0.193***	0.220***	0.187***	0.228***	0.256***	0.191***	0.119***	0.222***	0.285***	0.229***	0.279***	0.241***	0.117***	0.119***
	(0.022)	(0.022)	(0.022)	(0.024)	(0.028)	(0.025)	(0.024)	(0.027)	(0.030)	(0.024)	(0.028)	(0.030)	(0.028)	(0.026)	(0.030)
EEA	-0.000	0.000	0.000	-0.000	0.000	0.001*	0.001*	0.003**	0.002*	0.002*	-0.000	0.002**	0.001	0.000	0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
non-EEA	0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000	0.000	0.000	-0.000	-0.000	-0.000	-0.000	-0.001	-0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Age	0.034	0.139***	0.093**	0.100***	0.015	0.011	0.048*	0.025	0.043*	0.072***	0.063**	0.013	0.036*	0.059**	0.071**
	(0.032)	(0.031)	(0.032)	(0.030)	(0.023)	(0.018)	(0.022)	(0.021)	(0.020)	(0.021)	(0.020)	(0.016)	(0.016)	(0.019)	(0.026)
Age^2	-0.022	-0.081*	-0.019	-0.061	0.021	0.005	-0.024	-0.016	-0.038	-0.032	-0.048*	-0.003	-0.027	-0.050*	-0.031
	(0.032)	(0.032)	(0.033)	(0.031)	(0.023)	(0.018)	(0.022)	(0.022)	(0.020)	(0.020)	(0.019)	(0.016)	(0.016)	(0.019)	(0.025)
Experience	0.072**	-0.006	0.021	0.011	0.038*	0.041**	0.034*	0.045**	0.031*	0.003	0.013	0.039**	0.021	0.018	-0.003
	(0.022)	(0.021)	(0.022)	(0.020)	(0.017)	(0.014)	(0.016)	(0.016)	(0.015)	(0.014)	(0.013)	(0.013)	(0.011)	(0.013)	(0.017)
Experience^2	-0.078***	-0.055**	-0.093***	-0.051**	-0.065***	-0.046***	-0.049***	-0.044***	-0.027*	-0.038***	-0.019	-0.037***	-0.022*	-0.019	-0.034*
	(0.018)	(0.018)	(0.019)	(0.017)	(0.014)	(0.011)	(0.013)	(0.013)	(0.011)	(0.011)	(0.010)	(0.010)	(0.009)	(0.010)	(0.014)
Middle educ	-0.005**	-0.002	0.001	-0.001	-0.003	-0.003	-0.002	-0.005*	-0.002	-0.000	-0.002	-0.002	-0.002*	-0.003*	-0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
High educ	0.001	-0.001	-0.000	-0.001	-0.000	-0.001	-0.002	-0.003	-0.003*	-0.001	-0.003*	-0.009***	-0.007**	-0.008**	-0.004
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Years educ	0.029***	0.018*	0.016*	0.004	0.018*	0.022**	0.014*	-0.002	0.010	-0.001	0.005	-0.002	-0.010	-0.025*	-0.017*
	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.007)	(0.009)	(0.007)	(0.009)	(0.007)	(0.006)	(0.006)	(0.010)	(0.008)
Years educ2	-0.022***	-0.016*	-0.012	-0.000	-0.013*	-0.018*	-0.012	0.002	-0.008	0.000	-0.005	0.001	0.009	0.021*	0.014*
	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.006)	(0.008)	(0.006)	(0.008)	(0.007)	(0.006)	(0.006)	(0.009)	(0.007)
Training	0.002*	0.002	0.002*	0.003**	0.002*	0.001	-0.000	0.000	0.000	0.000	0.000	0.000	-0.000	-0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
In couple	-0.000	0.001	-0.001	-0.001	0.002	0.002	0.000	0.001	0.001	0.000	0.001	0.001	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
With dep. children	0.002*	0.001	0.000	0.002*	0.001	0.000	0.000	0.001	0.001	0.000	0.001	0.000	0.001	0.001	-0.000
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)
In couple with dep. children	0.002*	0.001	0.001	0.001	0.001	0.001	0.002	0.001	0.005**	0.002	0.002	0.001	0.002	0.001	0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Low segr.	0.001	0.007***	0.004*	0.002	0.002	0.004*	0.000	0.004**	0.003*	0.000	-0.001	-0.001	0.002	0.000	-0.004**
	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Public	-0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.001	0.001	0.001*	0.001*	0.001*	0.002*	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)
Coefficients															
ln(hours)	3.093***	3.009***	3.336***	3.982***	2.191***	2.002***	1.968***	1.393***	2.055***	4.865***	2.681***	3.365***	2.906***	3.063***	1.275*
	(0.395)	(0.358)	(0.367)	(0.393)	(0.430)	(0.346)	(0.403)	(0.389)	(0.437)	(0.477)	(0.415)	(0.448)	(0.497)	(0.482)	(0.565)
ln(hours)^2	-1.784***	-1.794***	-1.988***	-2.160***	-1.443***	-1.404***	-1.306***	-1.089***	-1.365***	-2.780***	-1.714***	-2.056***	-1.923***	-1.781***	-1.097***
	(0.204)	(0.192)	(0.197)	(0.210)	(0.227)	(0.189)	(0.216)	(0.212)	(0.234)	(0.248)	(0.225)	(0.242)	(0.262)	(0.256)	(0.310)
EEA	-0.002	0.000	0.000	-0.001	-0.000	-0.001	0.002	0.003	-0.001	-0.001	-0.007*	-0.001	-0.001	-0.003	0.001
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
non-EEA	-0.000	-0.001	0.000	-0.000	-0.000	-0.005*	-0.000	-0.004	0.000	-0.005	-0.001	0.002	-0.001	-0.004	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Age	-0.325	-1.068*	-0.739	-0.461	0.566	0.292	-0.397	0.349	-0.542	-0.620	-0.308	0.892	-0.224	-0.652	-0.117
	(0.515)	(0.458)	(0.489)	(0.542)	(0.581)	(0.546)	(0.564)	(0.572)	(0.587)	(0.588)	(0.573)	(0.576)	(0.587)	(0.589)	(0.695)
Age^2	0.174	0.373	0.179	0.110	-0.332	-0.114	0.224	-0.060	0.371	0.131	0.427	-0.342	0.248	0.544	-0.102
	(0.248)	(0.223)	(0.237)	(0.266)	(0.290)	(0.269)	(0.277)	(0.282)	(0.290)	(0.297)	(0.291)	(0.293)	(0.304)	(0.305)	(0.359)
Experience	0.024	0.348*	0.213	0.160	-0.181	-0.054	0.077	-0.157	0.117	0.223	-0.082	-0.315	-0.002	-0.009	0.211
	(0.187)	(0.163)	(0.176)	(0.196)	(0.218)	(0.201)	(0.206)	(0.200)	(0.213)	(0.203)	(0.194)	(0.194)	(0.196)	(0.196)	(0.234)
Experience^2	0.020	-0.047	0.034	0.030	0.170	0.056	-0.023	0.050	-0.084	0.029	-0.085	0.125	-0.060	-0.083	0.010
	(0.080)	(0.070)	(0.075)	(0.084)	(0.096)	(0.087)	(0.088)	(0.089)	(0.093)	(0.094)	(0.092)	(0.091)	(0.096)	(0.096)	(0.114)
Middle educ	0.004	0.004	0.013	0.005	-0.002	0.006	0.000	-0.012	0.004	0.012	-0.013	-0.008	-0.015	-0.015	-0.004
	(0.007)	(0.007)	(0.007)	(0.008)	(0.009)	(0.008)	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)	(0.009)	(0.010)	(0.010)	(0.011)
High educ	0.002	0.014	0.014	0.007	0.001	-0.001	-0.009	-0.018	-0.000	0.020	-0.022	-0.034*	-0.027	-0.031	-0.013
	(0.009)	(0.009)	(0.010)	(0.011)	(0.012)	(0.012)	(0.013)	(0.013)	(0.015)	(0.015)	(0.015)	(0.016)	(0.017)	(0.018)	(0.022)
Years educ	-0.169	-0.108	-0.040	0.254	0.186	0.616	0.516	0.157	0.721	0.360	0.770	0.885	0.731	0.502	1.203*
	(0.316)	(0.313)	(0.330)	(0.357)	(0.394)	(0.385)	(0.403)	(0.415)	(0.430)	(0.449)	(0.457)	(0.457)	(0.481)	(0.500)	(0.559)
Years educ2	0.073	0.062	0.042	-0.062	-0.059	-0.228	-0.229	-0.057	-0.311	-0.106	-0.321	-0.352	-0.259	-0.169	-0.523
	(0.141)	(0.141)	(0.149)	(0.163)	(0.179)	(0.176)	(0.186)	(0.193)	(0.199)	(0.210)	(0.215)	(0.215)	(0.228)	(0.238)	(0.269)
Training	-0.012	-0.011	-0.003	-0.006	-0.009	-0.007	0.002	0.001	0.003	-0.002	0.004	0.003	0.001	0.000	-0.003
	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
With dep. children	0.049***	0.037***	0.044***	0.050***	0.045***	0.028**	0.052***	0.043***	0.052***	0.030*	0.036**	0.026*	0.036**	0.033*	0.028*
	(0.010)	(0.010)	(0.010)	(0.011)	(0.012)	(0.011)	(0.012)	(0.012)	(0.012)	(0.012)	(0.013)	(0.013)	(0.013)	(0.013)	(0.014)
In couple with dep. children	-0.014	-0.005	-0.006	-0.008	-0.007	-0.002	-0.036***	-0.013	-0.030**	-0.020	-0.011	-0.008	-0.012	-0.014	-0.014
	(0.009)	(0.008)	(0.009)	(0.010)	(0.010)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)	(0.011)	(0.012)	(0.012)
Low segr.	-0.005	0.010*	0.002	0.002	0.005	0.010*	0.001	0.005	0.007	0.003	0.001	-0.001	0.013**	0.011*	0.001
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)
Public	-0.000	0.002	-0.001	0.001	-0.000	-0.001	-0.002	-0.004	-0.003	-0.003	-0.002	0.001	0.002	0.005*	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
	(0.333)	(0.303)	(0.319)	(0.339)	(0.371)	(0.335)	(0.356)	(0.359)	(0.383)	(0.401)	(0.380)	(0.387)	(0.409)	(0.416)	(0.464)
Observations	17,046	17,745	16,772	15,574	13,410	13,840	13,204	13,218	13,606	12,994	12,447	12,874	12,085	11,588	10,995

*Notes: Contribution of main socio-demographic characteristics, human capital attributes and sectoral indicators. Significance levels: pvalue<0.01 ***, pvalue<0.05 **, pvalue<0.1 *.*

Table B.19. *Selected covariates by* LASSO

Dep.var.	ln_wage		Permanent		Part-time work		ln_hrs		Remote work		Female dom. sector	
	Lasso	Post-est OLS	Lasso	Post-est OLS	Lasso	Post-est OLS	Lasso	Post-est OLS	Lasso	Post-est OLS	Lasso	Post-est OLS
Selected covariates												
female			0.048	0.047	-0.039	-0.040					0.118	0.115
incouple		0.024	0.043			0.008	0.011					
kids	0.019	0.019	0.045	0.032								
female#incouple												
0 1	0.081	0.081	0.010	0.024	-0.014	-0.011	0.025	0.024	0.002	-0.003	-0.012	-0.017
1 1			0.111	0.121	-0.108	-0.110						
female#kids												
0 1	0.038	0.035		-0.031	-0.028	0.026	0.027					
1 1			0.119	0.123	-0.106	-0.105						
1 0	-0.040	-0.039				-0.012	-0.014					
incouple#kids												
0 1		0.023	0.048	0.044 0.051	-0.028	-0.030						
1 0	-0.011	-0.014										
1 1	0.021	0.020				0.007	0.009					
age	0.010	0.008		-0.015	-0.033	0.038	0.048				-0.000	-0.001
c.age#c.age			0.000	0.000	-0.000	-0.000	0.000	0.000				
group2												
EEA	-0.020	-0.019	-0.008	-0.020	-0.077	-0.076	0.114	0.112				
NEEA	-0.035	-0.033	-0.016	-0.029	0.027	0.034	-0.017	-0.018	-0.004	-0.012	0.012	0.035
black	-0.052	-0.053		-0.036	-0.044	0.045	0.046					
asian	-0.051	-0.054		-0.006	-0.017	-0.023	-0.025					
other_ethn	-0.051	-0.053	-0.043	-0.059	0.028	0.028	-0.019	-0.021			0.037	0.064
muslim	-0.070	-0.068		0.093	0.103	-0.113	-0.115				0.063	0.095
crist	-0.046	-0.046		-0.007	-0.009							
other relig	0.015	0.019		0.013	0.016	-0.020	-0.020					
education												
intermediate educ	0.051	0.051	-0.023	-0.044	0.010	0.005	-0.022	-0.019			0.004	0.012
higher educ	0.120	0.121	-0.047	-0.075					0.005	0.007	-0.010	-0.027
yrseduc	0.094	0.128				0.001	0.001					
yrseduc2	-0.003	-0.004										
experience	0.013	0.015		-0.005	-0.000	0.002	-0.002					
experience2	-0.000	-0.000	-0.000	-0.000	0.000	-0.000	0.000	0.000				
trnopp	-0.006	-0.006	0.061	0.070	-0.0328	-0.034	0.057	0.058				
public	-0.037	-0.039	-0.155	-0.166	0.067	0.071	-0.138	-0.138	-0.040	-0.052	0.332	0.350
soc												
Professional O..	0.050	0.034			-0.087	-0.091	0.006	-0.028				
Associate Prof..	-0.164	-0.182			0.018	0.037	-0.108	-0.116	0.032	-0.000	-0.029	-0.037
Administrative.	-0.440	-0.467	-0.075	-0.090	0.095	0.105	-0.184	-0.194	-0.004	-0.049	-0.051	-0.062
Skilled Trades..	-0.436	-0.455			0.002	0.000	-0.015	-0.063				
Caring, Leisur.	-0.568	-0.597			0.169	0.175	-0.214	-0.218			0.307	0.335
Sales And Cust.	-0.542	-0.558			0.260	0.264	-0.253	-0.255	-0.0359	-0.082	0.309	0.337
Process, Plant..	-0.534	-0.550	-0.068	-0.095	0.044	0.057			-0.039	-0.088	-0.097	-0.104
Elementary Occ.	-0.592	-0.607	-0.141	-0.182	0.205	0.211	-0.209	-0.211	-0.038	-0.083	0.045	0.057
1. female#soc												
Managers, Dire.	-0.061	-0.084	0.0437	0.066	-0.115	-0.109	0.0428	0.040	0.006	-0.022		
Professional O..	-0.047	-0.054	-0.036	-0.049			0.003	0.011	0.045	0.088		
Associate Prof..	-0.029	-0.032			-0.024	-0.036	0.049	0.052				
Administrative.	0.039	0.048			0.047	0.054						
Skilled Trades..	-0.100	-0.102			0.158	0.176	-0.170	-0.175			0.135	0.227
Caring, Leisur.	0.035	0.048			0.017	0.020	-0.009	-0.009	-0.006	-0.051		
Sales And Cust.	-0.003	-0.006			0.099	0.102	-0.102	-0.102				
Process, Plant..	0.008	0.013	-0.055	-0.074		-0.087	-0.093		-0.051	-0.111		
Elementary Occ.			0.0903	0.136	0.170	0.173	-0.318	-0.319			0.089	0.105
benefit	-0.114	-0.114			0.193	0.193	-0.247	-0.247			0.055	0.063
wrkregion2												
Rest of Northe.	0.022	0.024	0.006	0.027	-0.022	-0.028	0.043	0.047				
South Yorkshire			-0.007	-0.012	0.027	0.030						
West Yorkshire	0.043	0.046			-0.021	-0.027	0.014	0.018				
Rest of Yorks	0.158	0.159	0.007	0.029	-0.009	-0.014	0.029	0.033				
East Midlands	0.004	0.006			-0.002	-0.007	0.008	0.010	0.004	0.010		
East of England	0.010	0.012			0.004	0.004	-0.015	-0.013				
Greater London	0.127	0.128			-0.013	-0.016	0.016	0.018	-0.016	-0.025	-0.020	-0.043
Rest of South	-0.003	-0.003			0.014	0.015	-0.006	-0.005				
South West	-0.074	-0.075	-0.019	-0.036	0.045	0.046	-0.054	-0.052	0.003	0.012	0.017	0.052
West Midland			0.009	0.030	0.025	0.027	0.014	0.018				
Rest of West M	-0.042	-0.043			-0.026	-0.025						
Greater Manche	-0.037	-0.038	-0.009	-0.036			-0.012	-0.011	-0.006	-0.018		
Merseyside	0.014	0.017			0.000	0.002	0.0139	0.017				
Rest of North .					-0.017	-0.016	0.007	0.016				
Wales	-0.083	-0.085			-0.025	-0.025						
Scotland	-0.069	-0.070							-0.008	-0.017	0.006	0.041
Nothern Ireland	-0.110	-0.114			-0.005	-0.018	0.029	0.033	-0.007	-0.023		
Outside UK	-0.044	-0.059	-0.248	-0.377			0.195	0.209				
$\lambda(BIC)$	13.581		146.395		30.862		8.590		75.061		264.217	

Note: The estimated models correspond to those with minimum BIC.

Table B.20. *Probit for female dominance, pooled sample*

	Coeff.	Std. error
Dep. var: Female dominance		
Woman	1.024***	0.045
Woman in couple	-0.146***	0.038
Woman w/t dep. children	-0.149***	0.040
In couple	-0.059	0.033
Dep. children	0.117**	0.040
In couple w/dep. children	-0.071	0.042
EEA	-0.059	0.033
Non-EEA	0.170***	0.032
Age	-0.026***	0.005
Age sqr.	0.000***	0.000
Higher educ.	-0.042	0.024
Years of educ.	0.052	0.033
Years of educ. sqr.	-0.001	0.001
SOC 3. Associate professional and technical occ.	-0.482***	0.042
SOC 4. Admin and secretarial occ.	-0.621***	0.038
SOC 5. Skilled trades	-0.035	0.093
SOC 6. Caring, leisure and other service	0.906***	0.049
SOC 7. Sales and customer service	0.253***	0.045
SOC 8. Process, plant and machine operatives	-1.226***	0.092
SOC 9. Elementary occ.	0.018	0.042
Man in SOC 3	0.345***	0.060
Man in SOC 4	0.484***	0.067
Man in SOC 5	-0.060	0.101
Man in SOC 6	0.286***	0.086
Man in SOC 7	0.433***	0.065
Man in SOC 8	0.789***	0.100
Man in SOC 9	0.099	0.053
Region Controls	Yes	
Observations	25,117	

*Notes: The estimates are use to calculate the propensity score used in the PSM. Robust standard errors. Significance levels: pvalue<0.01 ***, pvalue<0.05 **, pvalue<0.1 *.*

Table B.21. *ATT for permanent job*

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
Permanent	Unmatched	0.782	0.771	0.011	0.005	2.06
	ATT	0.782	0.771	0.011	0.008	1.30

Note: S.E. does not take into account that the propensity score is estimated. The matching method is single nearest-neighbour; five neighbors are used to calculate the matched outcome. Common support check: untreated units on support are 11,641; treated units on support are 13,424; treated units off support are 16.

Table B.22. *ATT for part-time job*

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
Part-time work	Unmatched	0.425	0.169	0.256	0.006	45.71
	ATT	0.424	0.289	0.135	0.008	17.03

Note: S.E. does not take into account that the propensity score is estimated. The matching method is single nearest-neighbour; five neighbors are used to calculate the matched outcome. Common support check: untreated units on support are 11,649; treated units on support are 13,444; treated units off support are 16.

Table B.23. *ATT for ln(hours)*

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
ln(hours)	Unmatched	3.288	3.547	-0.259	0.006	-39.96
	ATT	3.289	3.413	-0.125	0.009	-13.64

Note: S.E. does not take into account that the propensity score is estimated. The matching method is single nearest-neighbour; five neighbors are used to calculate the matched outcome. Common support check: untreated units on support are 11,654; treated units on support are 13,447; treated units off support are 16.

Table B.24. *ATT for remote work*

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
Remote work	Unmatched	0.034	0.071	-0.038	0.003	-13.49
	ATT	0.034	0.078	-0.044	0.005	-9.11

Note: S.E. does not take into account that the propensity score is estimated. The matching method is single nearest-neighbour; five neighbors are used to calculate the matched outcome. Common support check: untreated units on support are 11,654; treated units on support are 13,447; treated units off support are 16.

Table B.25. *ATT for ln(wage)*

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
ln(wage)	Unmatched	2.211	2.390	-0.180	0.007	-27.35
	ATT	2.211	2.305	-0.094	0.107	-8.81

Note: S.E. does not take into account that the propensity score is estimated. The matching method is single nearest-neighbour; five neighbors are used to calculate the matched outcome. Common support check: untreated units on support are 11,654; treated units on support are 13,447; treated units off support are 16.

Table B.26. *Covariate imbalance test, single components*

Variable	Mean Treated	Mean Control	%bias	t-test t	t-test p>|t|
Woman	0.65561	0.65589	-0.1	-0.05	0.961
Woman in couple	0.26749	0.26578	0.4	0.32	0.751
Woman w/t dep. children	0.25455	0.25113	0.9	0.65	0.519
In couple	0.39994	0.39725	0.5	0.45	0.652
Dep. children	0.34498	0.34072	0.9	0.74	0.461
In couple w/dep. children	0.2173	0.22093	-0.9	-0.72	0.472
EEA	0.06923	0.07237	-1.2	-1.00	0.316
Non-EEA	0.09831	0.10218	-1.3	-1.06	0.291
Age	34.751	34.191	4.4	3.65	0.000
Age sqr.	1366.7	1327.4	4.0	3.32	0.001
Higher educ.	0.27233	0.28165	-2.1	-1.71	0.088
Years of educ.	13.649	13.681	-1.1	-0.91	0.365
Years of educ. sqr.	194.79	195.48	-0.8	-0.67	0.505
SOC 3. Associate professional and technical occ.	0.09006	0.09098	-0.3	-0.26	0.792
SOC 4. Admin and secretarial occ.	0.09697	0.09568	0.4	0.36	0.719
SOC 5. Skilled trades	0.05332	0.05247	0.3	0.31	0.756
SOC 6. Caring, leisure and other service	0.18324	0.18267	0.2	0.12	0.905
SOC 7. Sales and customer service	0.15245	0.15272	-0.1	-0.06	0.951
SOC 8. Process, plant and machine operatives	0.03019	0.03043	-0.1	-0.11	0.909
SOC 9. Elementary occ.	0.18896	0.19285	-1.0	-0.81	0.418
Man in SOC 3	0.05585	0.0567	-0.4	-0.30	0.763
Man in SOC 4	0.08017	0.07896	0.4	0.37	0.715
Man in SOC 5	0.01101	0.0108	0.2	0.16	0.869
Man in SOC 6	0.1555	0.15606	-0.2	-0.13	0.898
Man in SOC 7	0.10597	0.10422	0.7	0.47	0.639
Man in SOC 8	0.00454	0.00491	-0.4	-0.44	0.657
Man in SOC 9	0.11519	0.1194	-1.5	-1.07	0.283
Rest of Northern region	0.05466	0.05618	-0.7	-0.54	0.587
South Yorkshire	0.04298	0.04518	-1.1	-0.88	0.379
West Yorkshire	0.05444	0.0538	0.3	0.23	0.817
Rest of York & Humberside	0.05748	0.05357	1.6	1.40	0.161
East Midlands	0.09727	0.09906	-0.6	-0.49	0.623
East of England	0.08314	0.08069	0.9	0.73	0.463
Greater London	0.08113	0.08574	-1.6	-1.37	0.172
Rest of South East	0.10248	0.10854	-2.0	-1.62	0.105
South West	0.0647	0.06113	1.5	1.21	0.228
West Midlands	0.05577	0.05665	-0.4	-0.31	0.755
Rest of West Midlands	0.04083	0.03799	1.5	1.20	0.231
Greater Manchester	0.02774	0.02716	0.3	0.29	0.771
Merseyside	0.04127	0.03791	1.7	1.41	0.158
Rest of North West	0.06083	0.06574	-2.1	-1.65	0.098
Wales	0.03109	0.03109	-0.0	-0.00	1.000
Scotland	0.06031	0.05693	1.4	1.18	0.239
Nothern Ireland	0.01495	0.01321	1.5	1.21	0.226
Outside UK	0.00074	0.00101	-0.8	-0.74	0.458

*Notes: * if variance ratio outside [0.96; 1.04]. When* $|\%bias| < 5$*, the balancing property is satisfied (Rosenbaum and Rubin, 1985)*

Table B.27. *Covariate imbalance test*

Ps R^2	LR χ^2	p> χ^2	Mean Bias	Median Bias	B	R	%Var
0.001	51.40	0.237	1.0	0.8	8.7	1.03	0

*Note: * if B> 25%, R outside [0.5; 2].*

Table B.28. *Mincerian regression results, years 2005-2020*

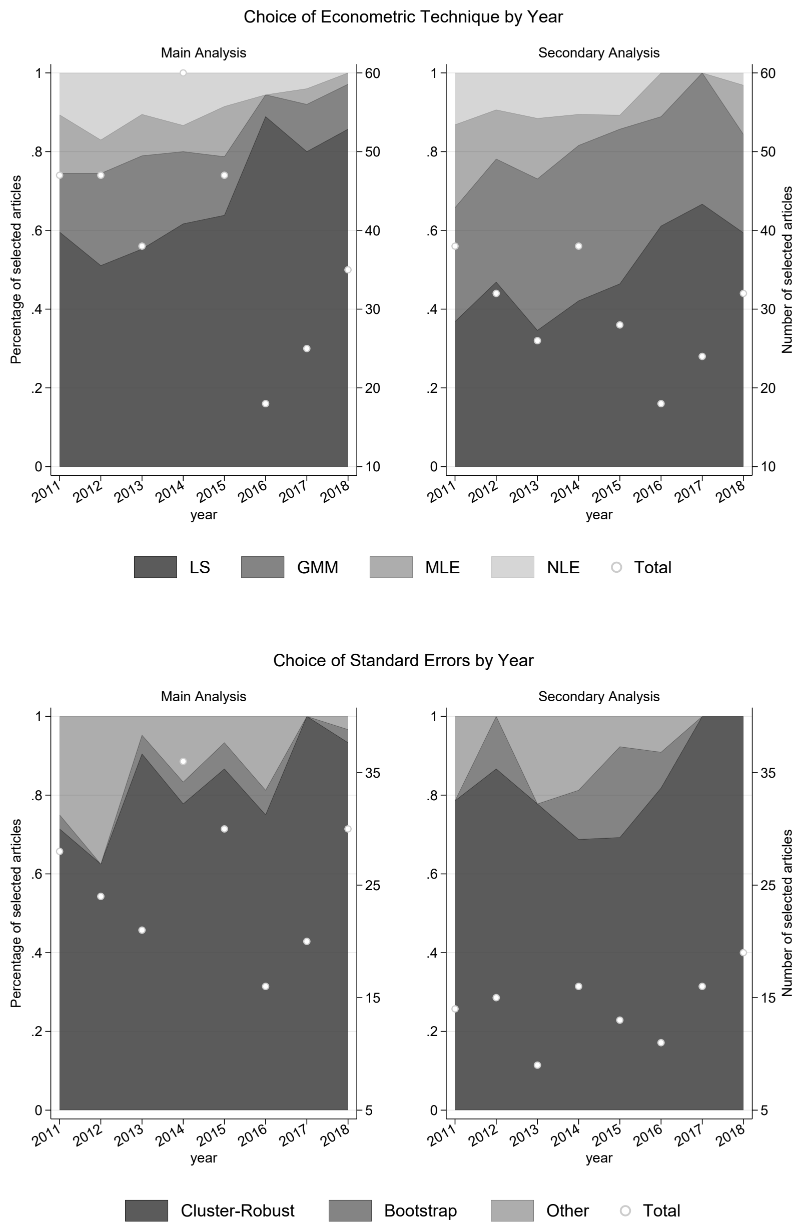
	Dep. var.: Log(Wage)					
	All sectors		Male-dominated sectors		Female-dominated sectors	
	Man (1)	Women (2)	Man (3)	Women (4)	Man (5)	Women (6)
Socio-demographic variables						
Age	0.007***	0.005***	0.009***	0.014***	0.005*	-0.000
	(0.001)	(0.001)	(0.002)	(0.003)	(0.002)	(0.001)
Age^2	0.000***	0.000***	0.000	-0.000	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
EEA	-0.056***	-0.044***	-0.062***	-0.064***	-0.038***	-0.020***
	(0.004)	(0.004)	(0.005)	(0.006)	(0.007)	(0.004)
Non-EEA	-0.032***	-0.017***	-0.009	-0.003	-0.063***	-0.028***
	(0.004)	(0.004)	(0.006)	(0.007)	(0.007)	(0.005)
In couple	0.064***	0.015***	0.064***	0.019***	0.061***	0.012***
	(0.002)	(0.002)	(0.003)	(0.004)	(0.004)	(0.003)
With dependent children	0.041***	-0.029***	0.042***	-0.034***	0.040***	-0.027***
	(0.004)	(0.003)	(0.005)	(0.005)	(0.006)	(0.003)
In couple with dep. children	-0.001	0.020***	-0.001	0.045***	-0.007	0.007
	(0.004)	(0.003)	(0.005)	(0.006)	(0.007)	(0.004)
Human capital variables						
Intermediate education	0.022***	0.021***	0.026***	0.028***	0.017***	0.016***
	(0.003)	(0.002)	(0.004)	(0.005)	(0.004)	(0.003)
High education	0.098***	0.101***	0.105***	0.106***	0.088***	0.093***
	(0.006)	(0.005)	(0.007)	(0.010)	(0.009)	(0.006)
Years of education	0.158***	0.117***	0.166***	0.135***	0.130***	0.104***
	(0.003)	(0.003)	(0.004)	(0.006)	(0.006)	(0.003)
Years of education^2	-0.005***	-0.003***	-0.005***	-0.004***	-0.004***	-0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Experience	0.015***	0.013***	0.017***	0.016***	0.013***	0.011***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Experience^2	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Training	0.068***	0.049***	0.075***	0.073***	0.051***	0.036***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)
Workplace characteristics						
Part-time	-0.097***	-0.038***	-0.103***	-0.044***	-0.099***	-0.037***
	(0.003)	(0.002)	(0.005)	(0.003)	(0.004)	(0.002)
Public sector	0.032***	0.059***	-0.020***	-0.038***	0.087***	0.083***
	(0.003)	(0.002)	(0.005)	(0.006)	(0.005)	(0.002)
Low gender segregation	0.027***	-0.005	0.024***	-0.008*	0.059***	0.034***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)
Female dominance	-0.163***	-0.158***				
	(0.003)	(0.002)				
Female Dominance ×	0.006	0.016***				
Low gender segregation	(0.005)	(0.004)				
Working region controls	Yes	Yes	Yes	Yes	Yes	Yes
Other demographic controls	Yes	Yes	Yes	Yes	Yes	Yes
Job controls	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	218,696	219,173	147,666	76,650	71,030	142,523

*Notes: Data from UK Labour Force Survey (LFS). Models (1)-(4) are estimated using OLS. Robust errors are in parenthesis. Significance levels: pvalue<0.01 ***, pvalue<0.05 **, pvalue<0.1 *.*

Appendix C

Figures

Figure C.1. *Articles in AER using panel data models*



NOTE: *The top graph gives a general picture of the choice of econometric techniques for linear panel data models – grouped in four comprehensive categories: LS, MLE, GMM, NLS – over the period from 2011 to 2018. On the left, the total number of selected articles are displayed for the* main *analysis by estimation method while, on the right, for the* secondary. *The majority of selected studies use linear least squares for panel data models, and this trend shows some regularities over the eight years in analysis. Despite the popularity of LS estimation, the choice of other estimation techniques remains limited touching the same figures in both primary and secondary analyses with the exception of GMM figures that tend to grow in the* secondary *plot. The bottom graph provides a detailed picture of the choice of standard errors by year of publication among articles that use least squares methods for linear panel data. The total number of eligible articles is displayed by standard error formula on the left for the* main *analysis while the on the right for the* secondary. *A growing number of those uses robust or cluster-robust standard errors in both* main *and* secondary *analyses whereas a decreasing minority makes use of bootstrapped standard errors or other formulae.*

Figure C.2. *Power test for* β_1*, heteroskedasticity*

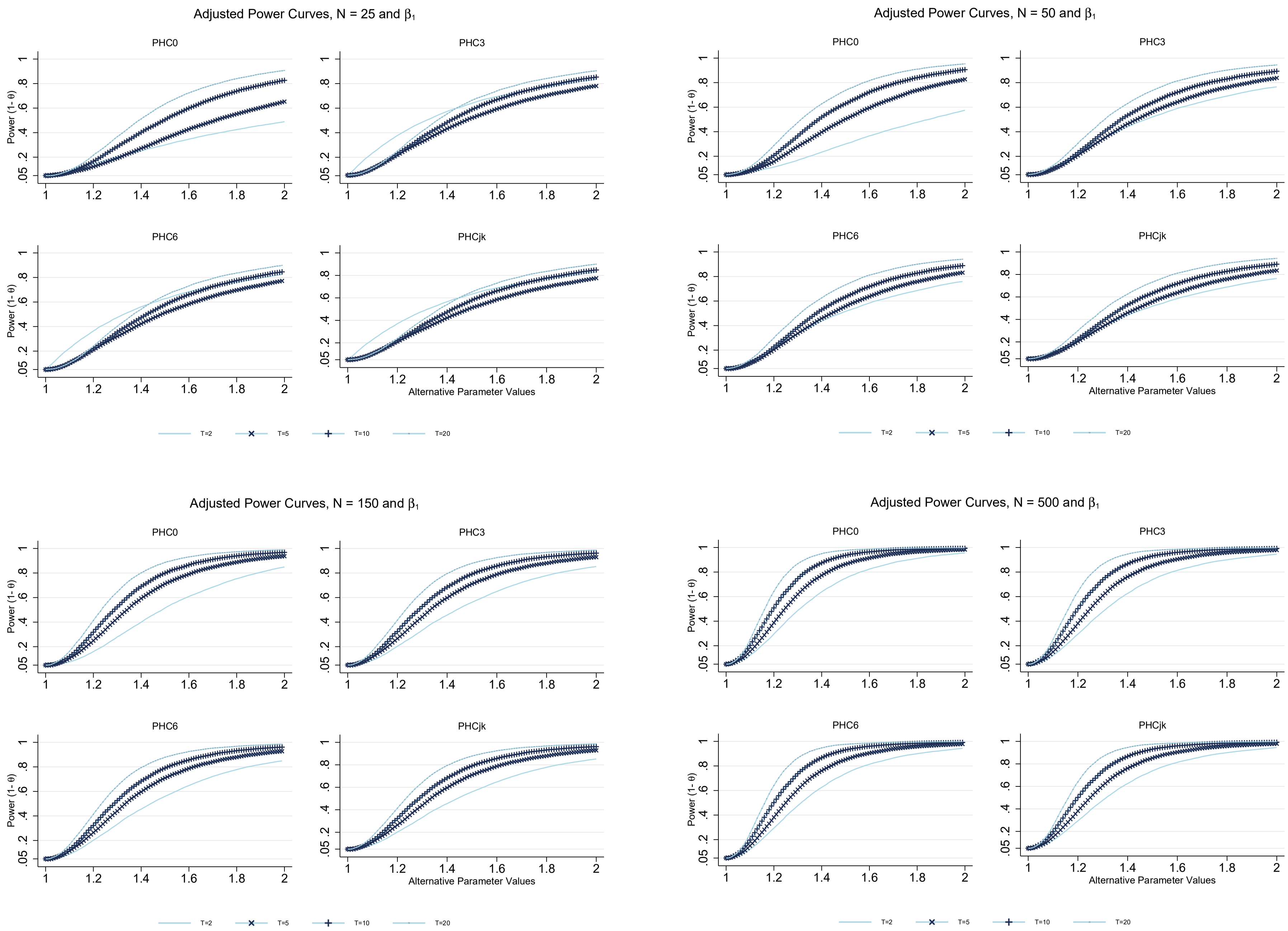


Figure C.3. *Power test for* β_1*, homoskedasticity*

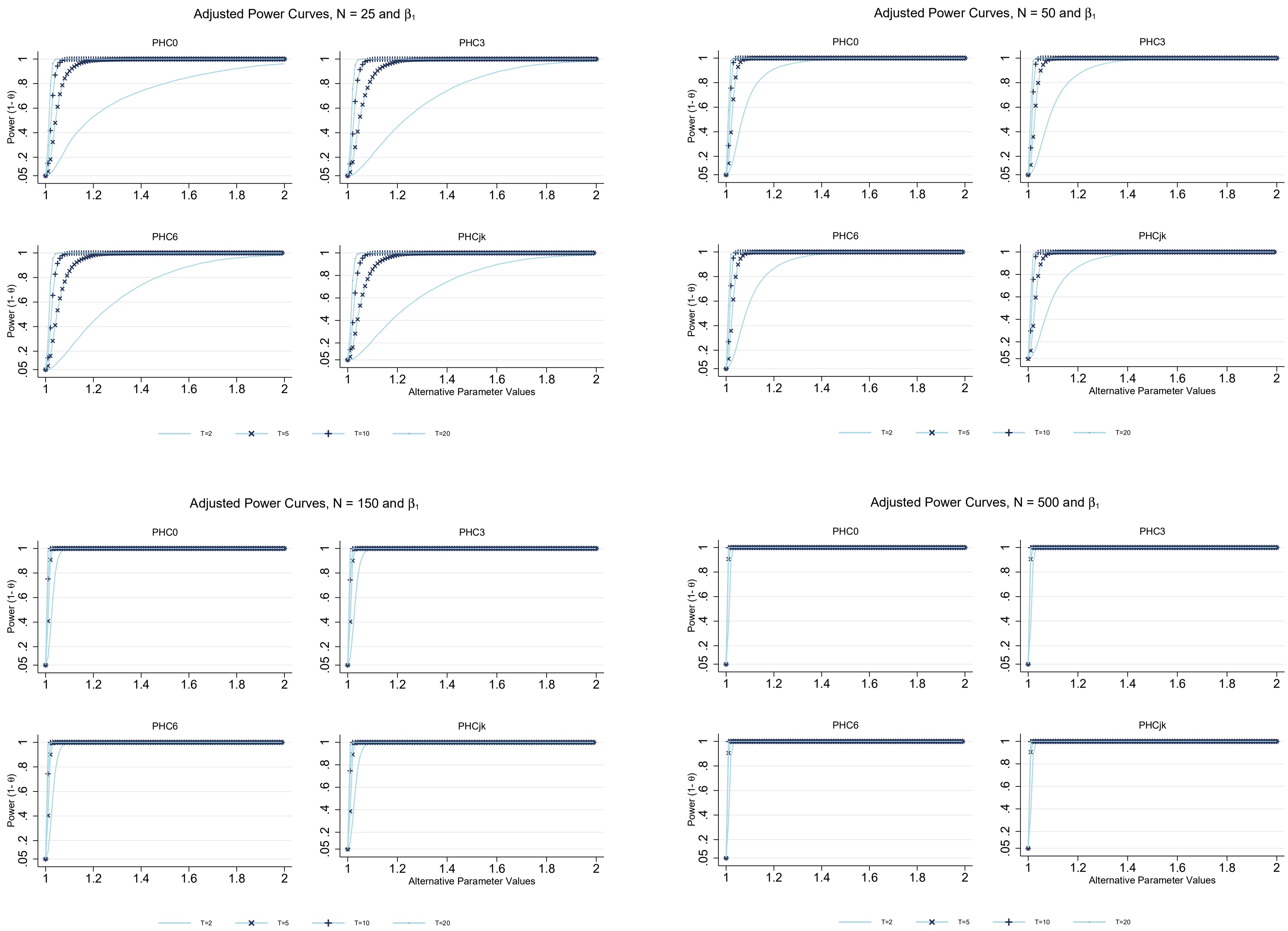


Figure C.4. *Power test for joint coefficient test, heteroskedasticity*

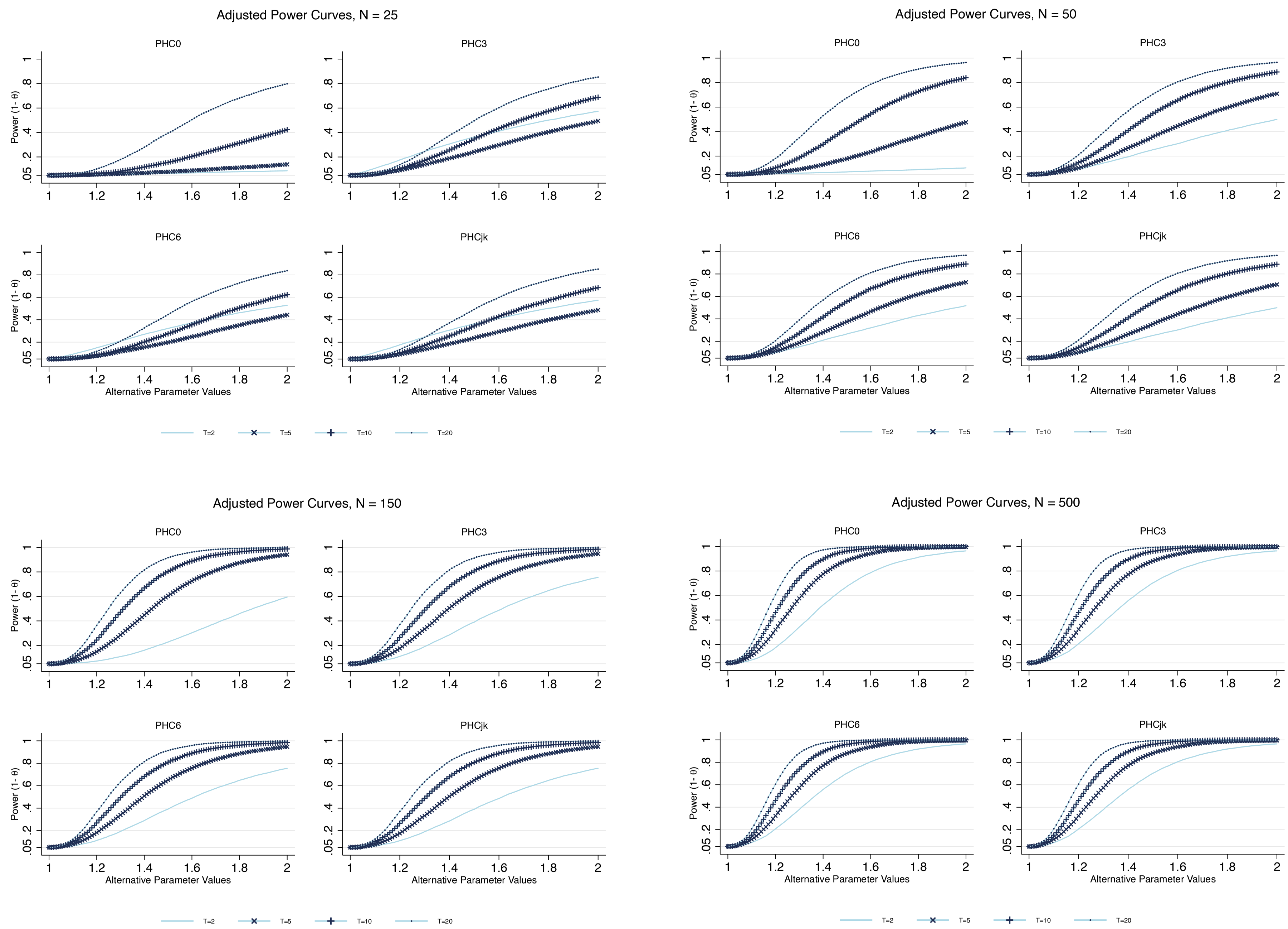


Figure C.5. *Power test for joint coefficient test, homoskedasticity*

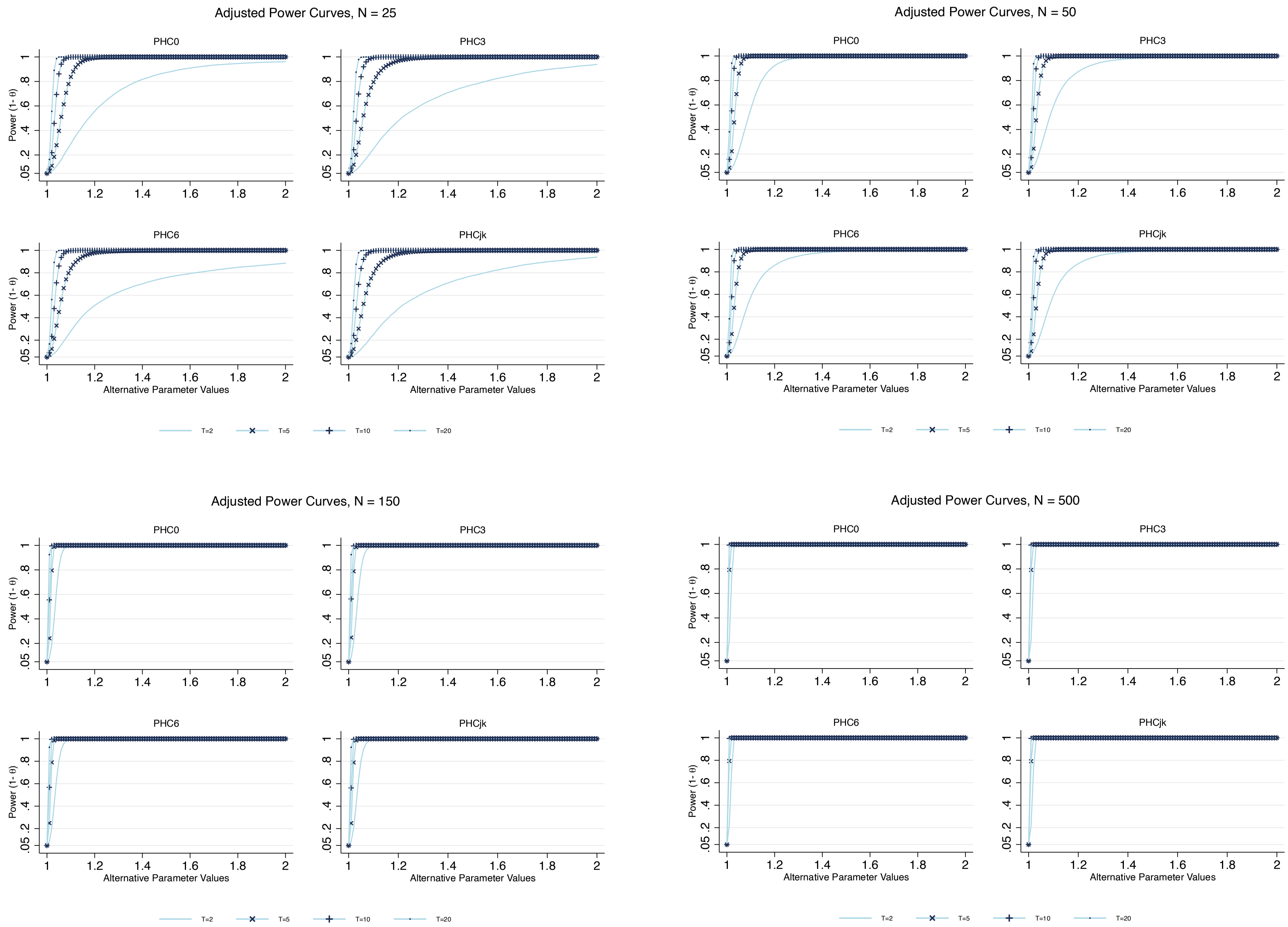
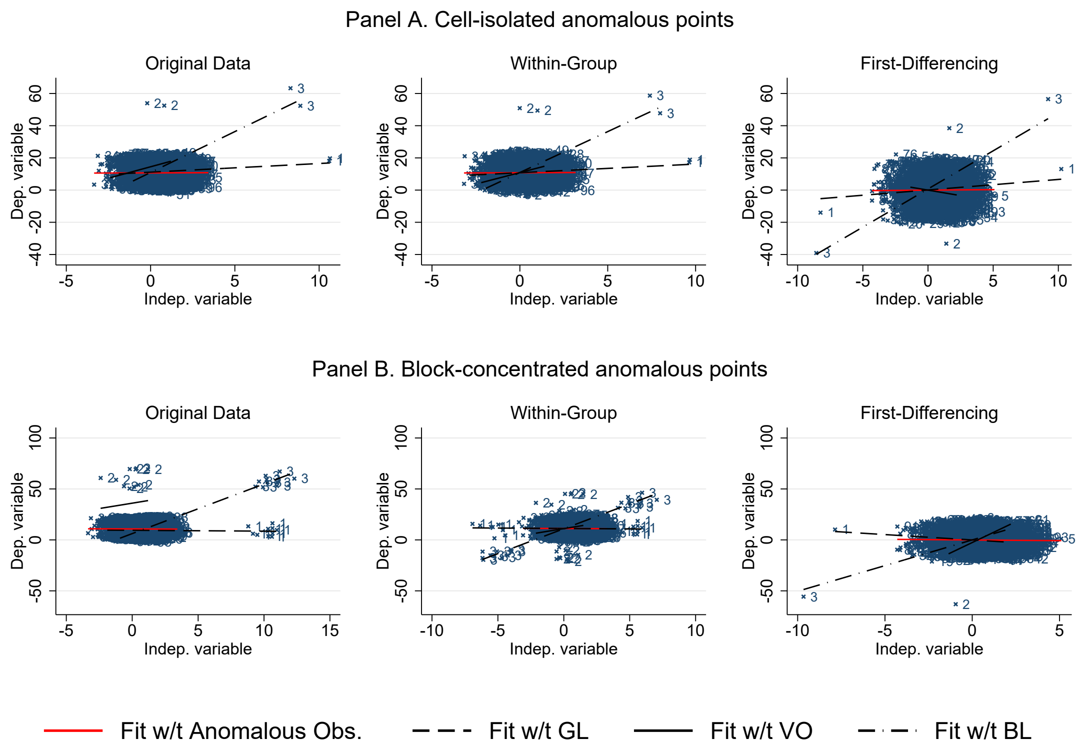


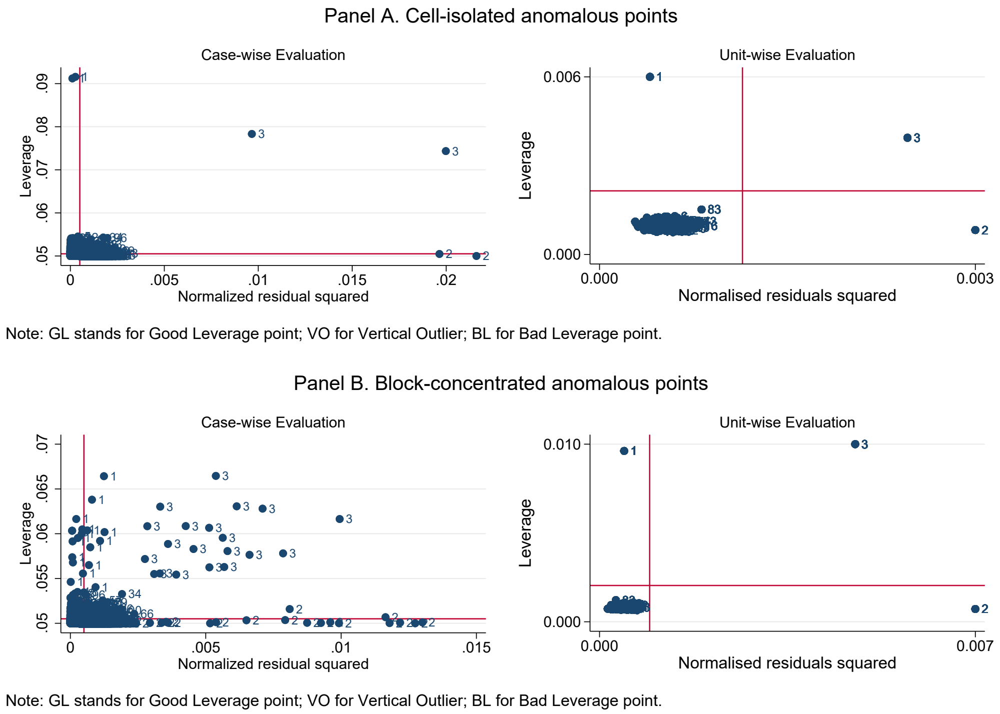
Figure C.6. *Example of anomalous points in panel data*



Note: GL stands for Good Leverage point; VO for Vertical Outlier; BL for Bad Leverage point.

From left to right: original data; data after the within-group transformation; and data after the first-differencing. Data is simulated following Bramati and Croux's (2007) design; here, N=100 and T=20. In both panels of scatter plots: unit 1 is a "good" leverage point; unit 2 is a vertical outlier; unit 3 is a "bad" leverage point. In Panel A, the first three units are contaminated at time $t = 1$ *and* $t = 20$*; in Panel B, half of each series for* $t \leq 10$ *is contaminated. In scatter plots, the red dotted line is fitted using uncontaminated units only; the dashed line using uncontaminated and good leverage points; the dash-dot line using uncontaminated and bad leverage points; the solid line uncontaminated units and vertical outliers.*

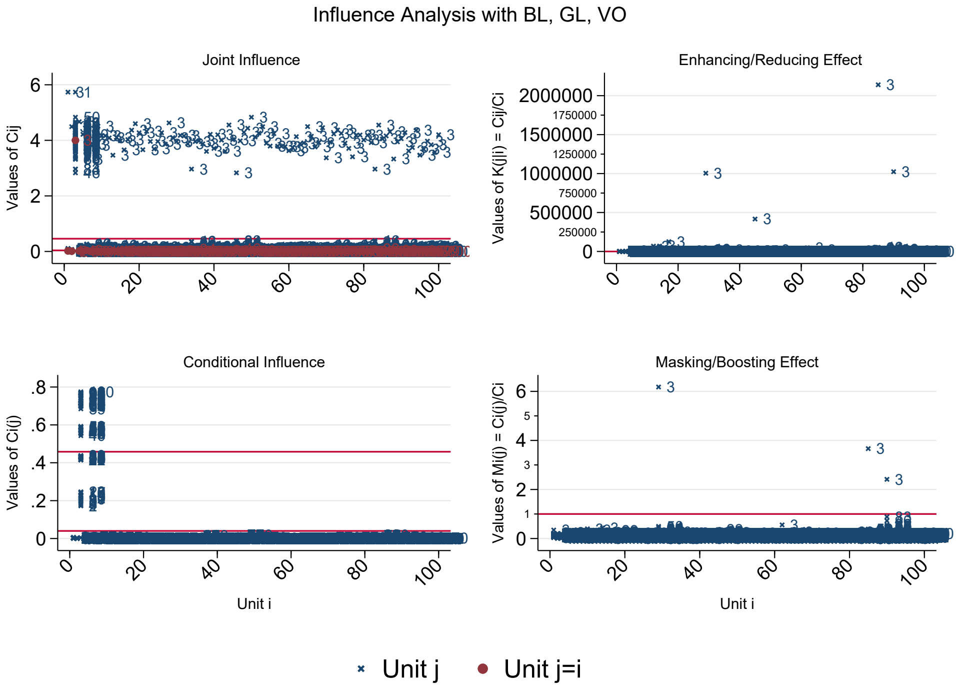
Figure C.7. *Leverage-residual plots*



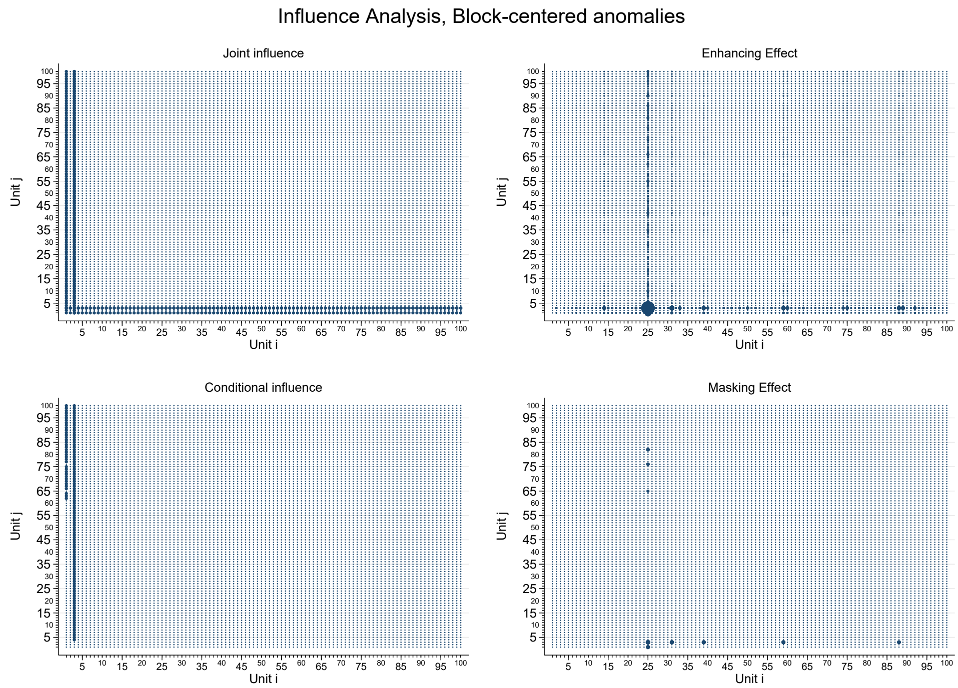
Leverage against normalised residual squared plot from assessing each case individually (left) and from assessing each unit in its whole history (right). Data is simulated following Bramati and Croux's (2007) design; here, N=100 and T=20. In both panels of scatter plots: unit 1 is a "good" leverage point; unit 2 is a vertical outlier; unit 3 is a "bad" leverage point. In Panel A, the first three units are contaminated at time $t = 1$ *and* $t = 20$*; in Panel B, half of each series for* $t \leq 10$ *is contaminated. The horizontal and vertical red lines are the means for leverage and for the normalized residual squared, respectively. Those points above the horizontal line display high leverage whereas points to the right of the vertical line have large residuals. Those points with both high leverage and large squared residuals are "bad" leverage points.*

Figure C.8. *Influence analysis with 2-way graphs (BL, GL, VO)*

(a) *Cell-isolated units*



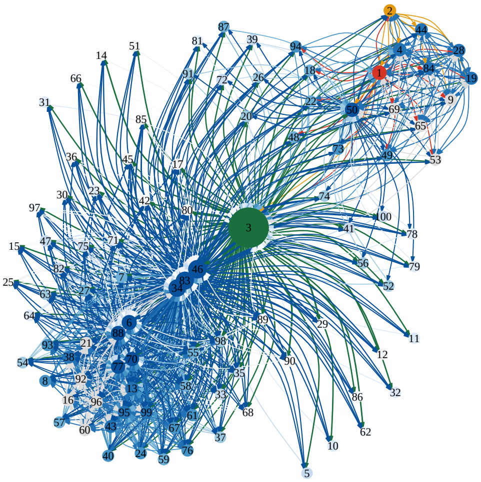
(b) *Block-centered*



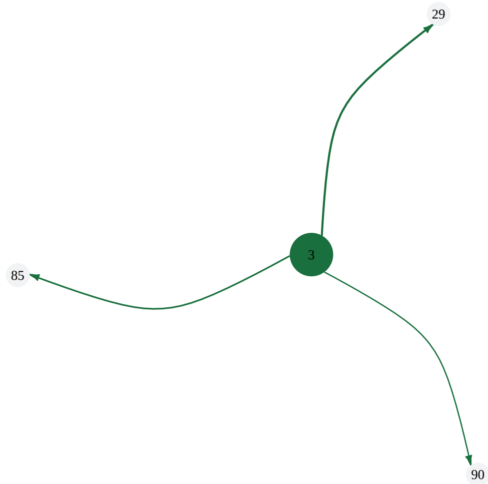
NOTE: *Data is simulated following Bramati and Croux's (2007) design; here, N=100 and T=20. In both panels of scatter plots: unit 1 is a "good" leverage point; unit 2 is a vertical outlier; unit 3 is a "bad" leverage point. In Panel (a), the first three units are contaminated at time* $t = 1$ *and* $t = 20$*; in Panel (b), half of each series for* $t \leq 10$ *is contaminated. The solid red lines display the cutoff values* $c_1 = F(k, N-k, 0.5)$ *and* $c_2 = 4/N$ *(in the left graphs) and* $c \geq 1$ *(in the right graphs).*

Figure C.9. *Influence analysis with network graphs (BL, GL, VO)*

Panel A. Cell−isolated anomalous units, all.

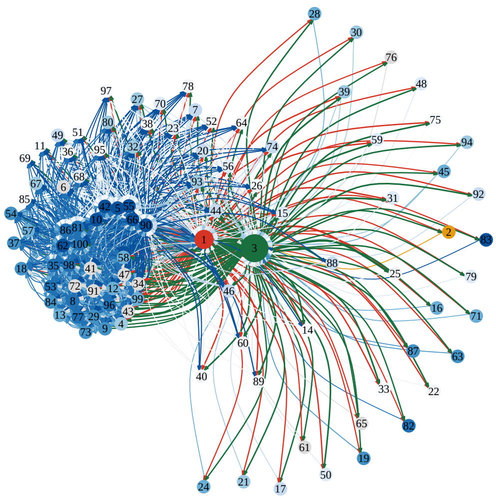


Enhancing effect

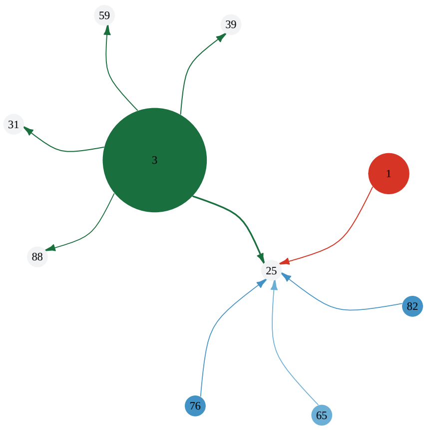


Masking effect

Panel B. Block−centered anomalous units, all.



Enhancing effect

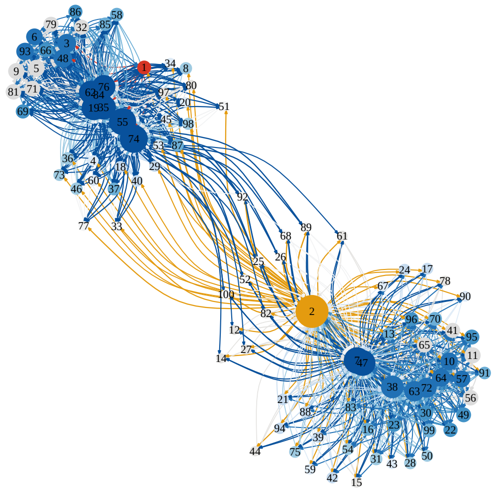


Masking effect

NOTE: *Unit 1 (in red) is a "good" leverage point; unit 2 (in orange) is a vertical outlier; unit 3 (in green) is a "bad" leverage point.* ENHANCING EFFECT*: Units whose effect is enhanced by unit j or enhance the effect of unit i and their joint influence exceeds the cut-off of 4/N.* MASKING EFFECT*: Units whose effect is masked by unit j or masks the effect of unit i based on the conditional influence. The size (from small to large) and colour (from light to dark blue) of the nodes reflect the degree of the total joint and conditional influence of unit i. The width of the links reflects the strength if the enhancing and masking effects; their colours match with the color of the nodes.*

Figure C.10. *Influence analysis with network graphs (two GL)*

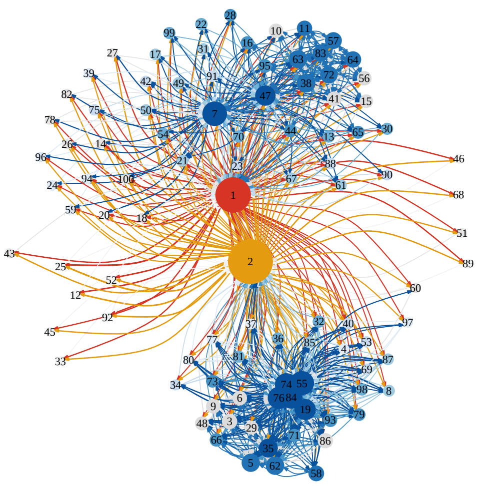
Panel A. Cell−isolated anomalous units, two GL.



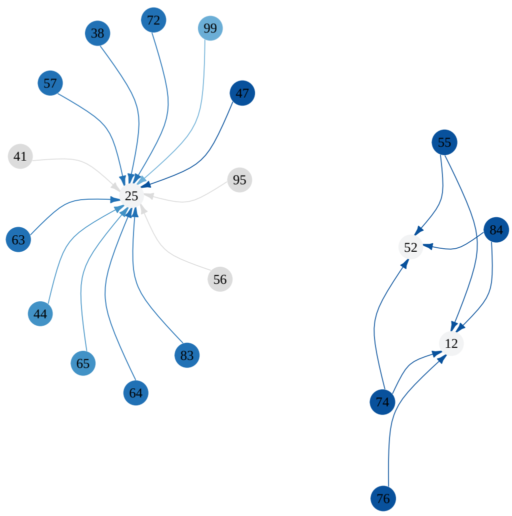
Enhancing effect

Masking effect

Panel B. Block−centered anomalous units, two GL.



Enhancing effect

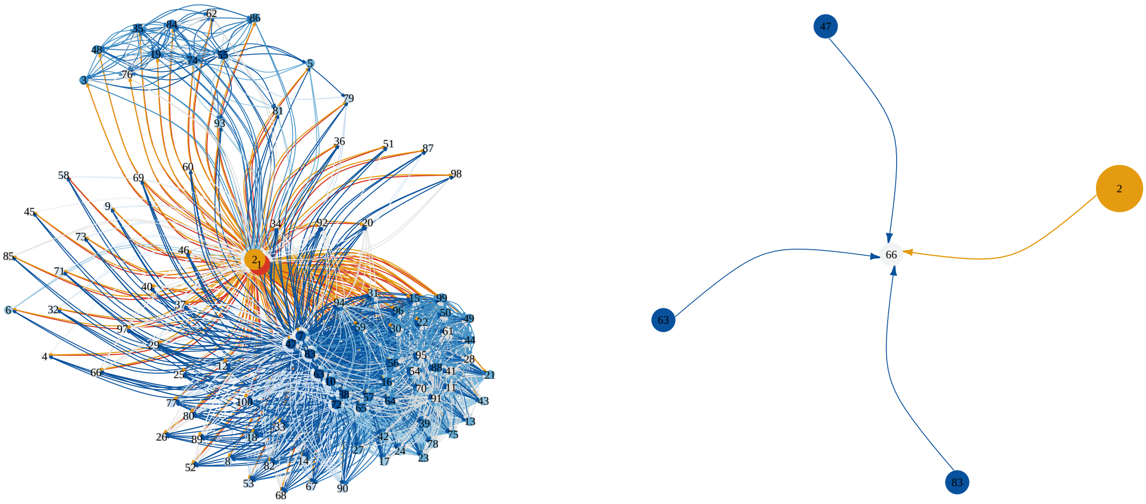


Masking effect

NOTE: *Unit 1 (in red) and unit 2 (in orange) are "good" leverage points.* ENHANCING EFFECT*: Units whose effect is enhanced by unit j or enhance the effect of unit i and their joint influence exceeds the cut-off of 4/N.* MASKING EFFECT*: Units whose effect is masked by unit j or masks the effect of unit i based on the conditional influence. The size (from small to large) and colour (from light to dark blue) of the nodes reflect the degree of the total joint and conditional influence of unit i. The width of the links reflects the strength if the enhancing and masking effects; their colours match with the color of the nodes.*

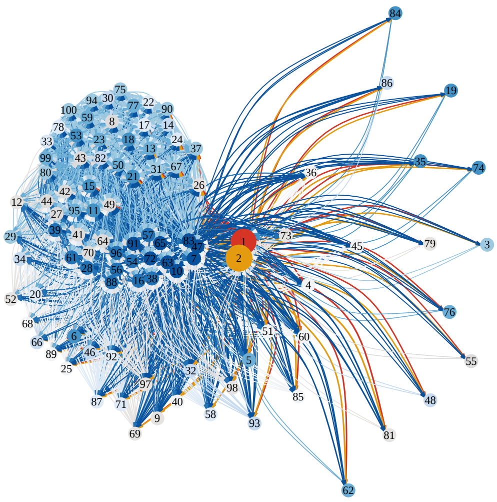
Figure C.11. *Influence analysis with network graphs (two BL)*

Panel A. Cell−isolated anomalous units, two BL.



Enhancing effect | Masking effect

Panel B. Block−centered anomalous units, two BL.

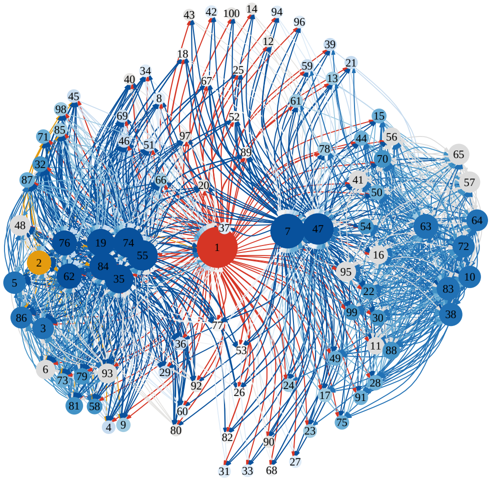


Enhancing effect | Masking effect

NOTE: *Unit 1 (in red) and unit 2 (in orange) are "bad" leverage points.* ENHANCING EFFECT*: Units whose effect is enhanced by unit j or enhance the effect of unit i and their joint influence exceeds the cut-off of 4/N.* MASKING EFFECT*: Units whose effect is masked by unit j or masks the effect of unit i based on the conditional influence. The size (from small to large) and colour (from light to dark blue) of the nodes reflect the degree of the total joint and conditional influence of unit i. The width of the links reflects the strength if the enhancing and masking effects; their colours match with the color of the nodes.*

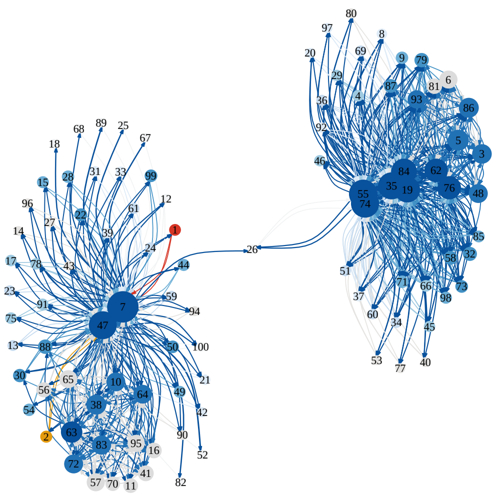
Figure C.12. *Influence analysis with network graphs (two VO)*

Panel A. Cell−isolated anomalous units, two VO.



Enhancing effect Masking effect

Panel B. Block−centered anomalous units, two VO

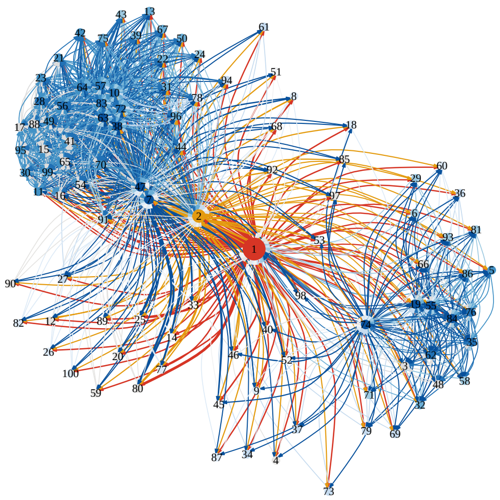


Enhancing effect Masking effect

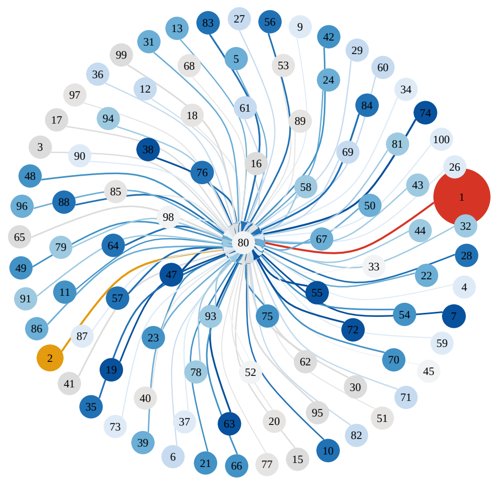
NOTE: *Unit 1 (in red) and unit 2 (in orange) are vertical outliers.* ENHANCING EFFECT*: Units whose effect is enhanced by unit j or enhance the effect of unit i and their joint influence exceeds the cut-off of 4/N.* MASKING EFFECT*: Units whose effect is masked by unit j or masks the effect of unit i based on the conditional influence. The size (from small to large) and colour (from light to dark blue) of the nodes reflect the degree of the total joint and conditional influence of unit i. The width of the links reflects the strength if the enhancing and masking effects; their colours match with the color of the nodes.*

Figure C.13. *Influence analysis with network graphs (BL, GL)*

Panel A. Cell−isolated anomalous units, BL and GL

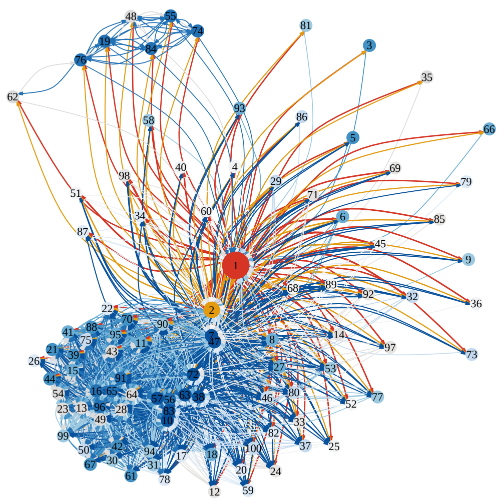


Enhancing effect

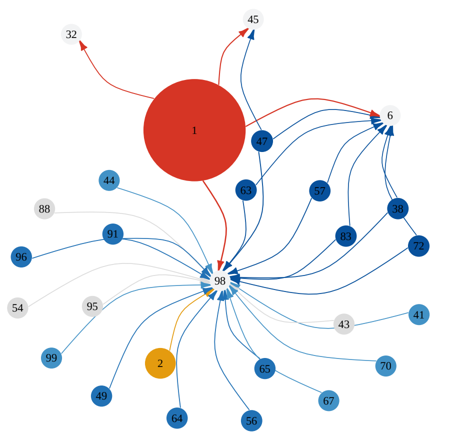


Masking effect

Panel B. Block−centered anomalous units, BL and GL



Enhancing effect

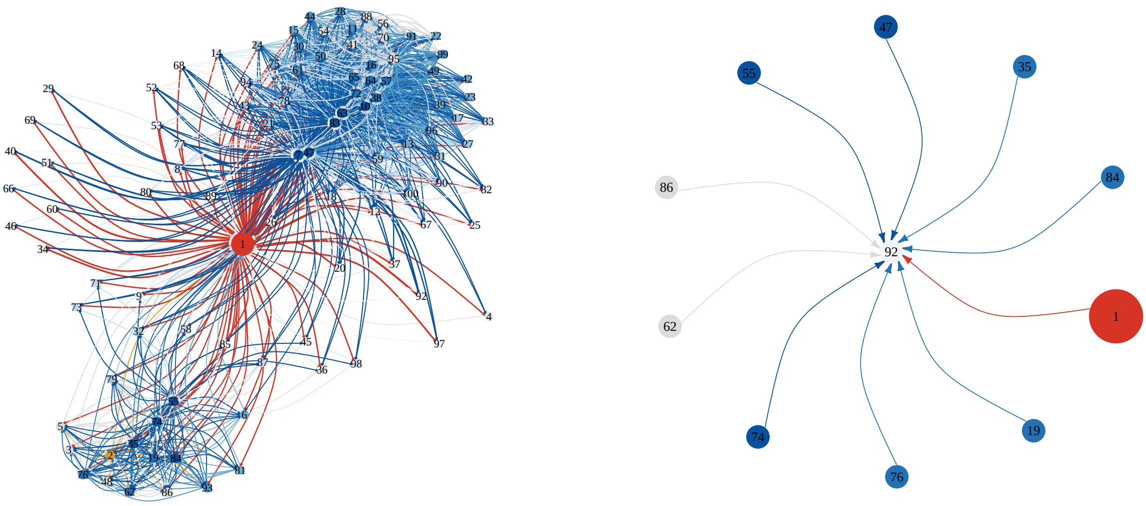


Masking effect

NOTE: *Unit 1 (in red) is a "bad" leverage point and unit 2 (in orange) a "good" leverage point.* ENHANCING EFFECT*: Units whose effect is enhanced by unit j or enhance the effect of unit i and their joint influence exceeds the cut-off of 4/N.* MASKING EFFECT*: Units whose effect is masked by unit j or masks the effect of unit i based on the conditional influence. The size (from small to large) and colour (from light to dark blue) of the nodes reflect the degree of the total joint and conditional influence of unit i. The width of the links reflects the strength if the enhancing and masking effects; their colours match with the color of the nodes.*

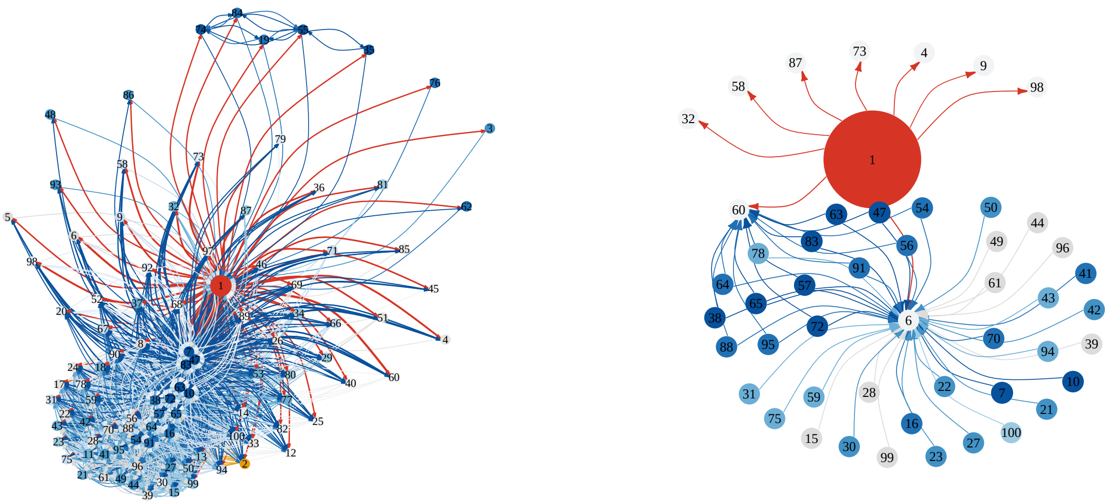
Figure C.14. *Influence analysis with network graphs (BL, VO)*

Panel A. Cell−isolated anomalous units, BL and VO.



Enhancing effect — Masking effect

Panel B. Block−centered anomalous units, BL and VO.

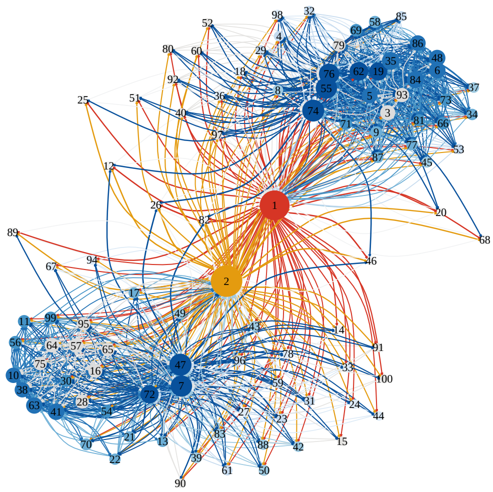


Enhancing effect — Masking effect

NOTE: *Unit 1 (in red) is a "bad" leverage point and unit 2 (in orange) a vertical outlier.* ENHANCING EFFECT*: Units whose effect is enhanced by unit j or enhance the effect of unit i and their joint influence exceeds the cut-off of 4/N.* MASKING EFFECT*: Units whose effect is masked by unit j or masks the effect of unit i based on the conditional influence. The size (from small to large) and colour (from light to dark blue) of the nodes reflect the degree of the total joint and conditional influence of unit i. The width of the links reflects the strength if the enhancing and masking effects; their colours match with the color of the nodes.*

Figure C.15. *Influence analysis with network graphs (GL, VO)*

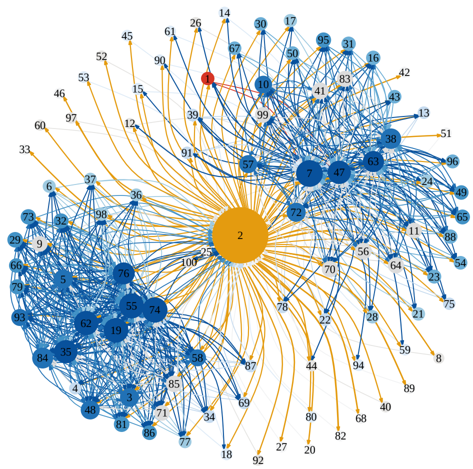
Panel A. Cell−isolated anomalous units, VO and GL



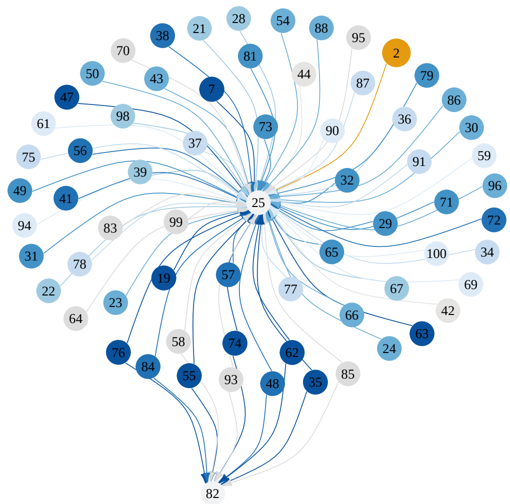
Enhancing effect

Masking effect

Panel B. Block−centered anomalous units, VO and GL.



Enhancing effect

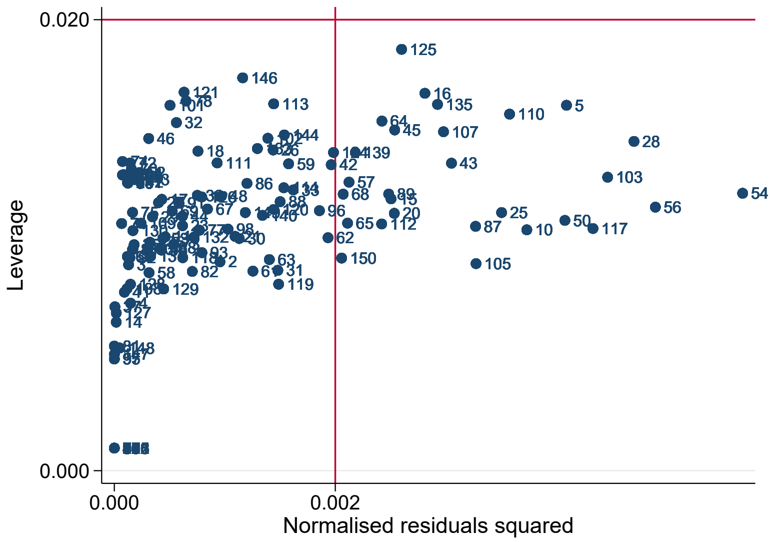


Masking effect

NOTE: *Unit 1 (in red) is a vertical outlier and unit 2 (in orange) a "good" leverage point.* ENHANCING EFFECT*: Units whose effect is enhanced by unit j or enhance the effect of unit i and their joint influence exceeds the cut-off of 4/N.* MASKING EFFECT*: Units whose effect is masked by unit j or masks the effect of unit i based on the conditional influence. The size (from small to large) and colour (from light to dark blue) of the nodes reflect the degree of the total joint and conditional influence of unit i. The width of the links reflects the strength if the enhancing and masking effects; their colours match with the color of the nodes.*

Figure C.16. *Leverage-Residual plot for Acemoglu et al. (2008)*

(a) *Model (2)*



(b) *Model (7)*

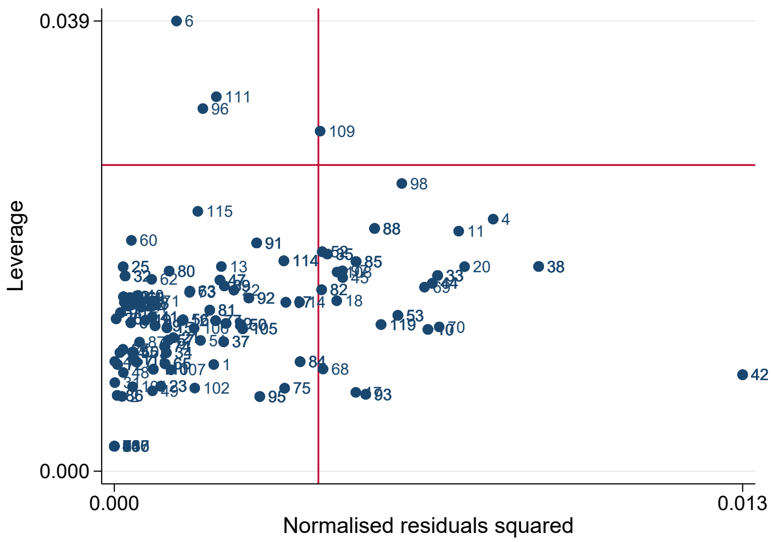
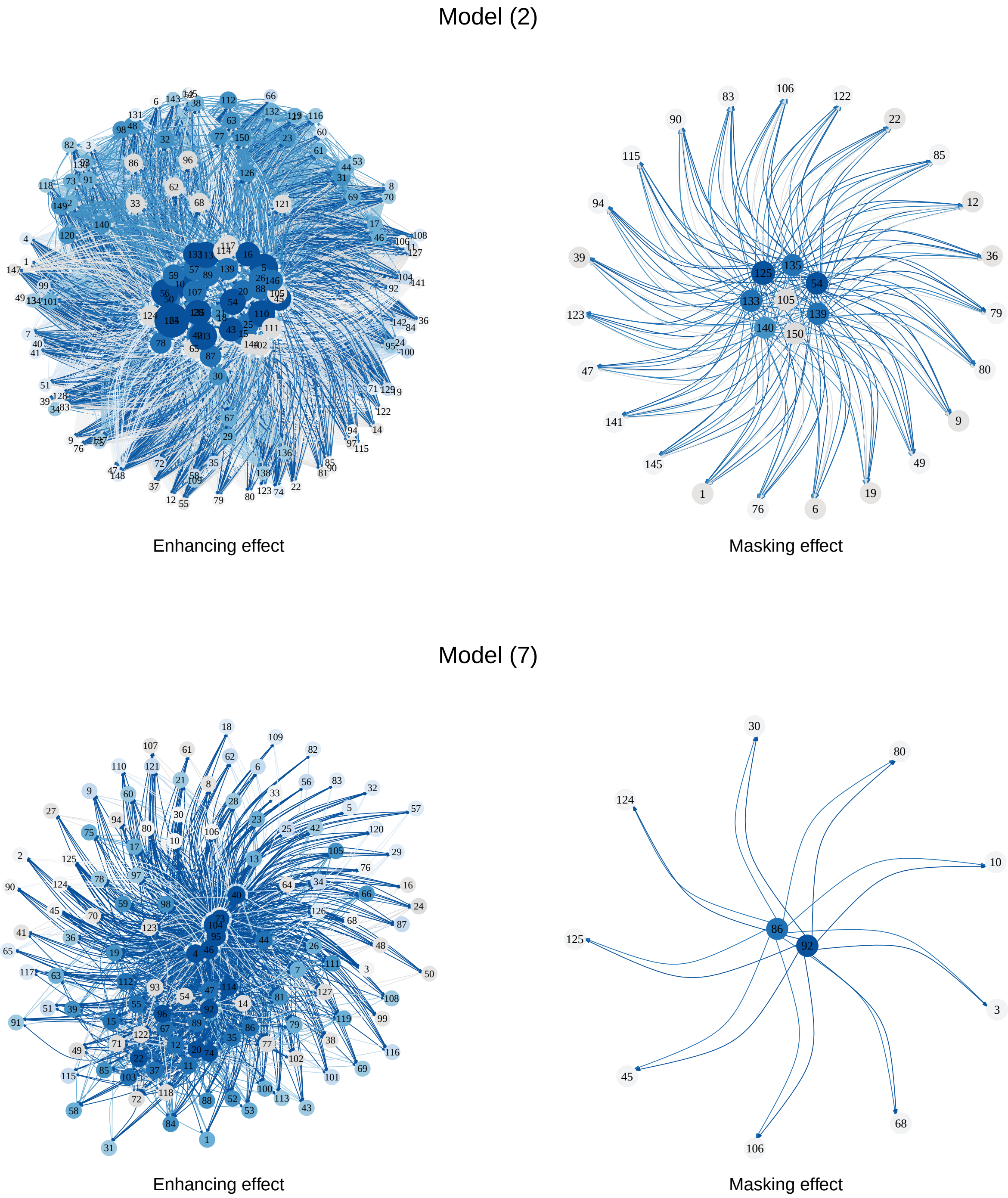


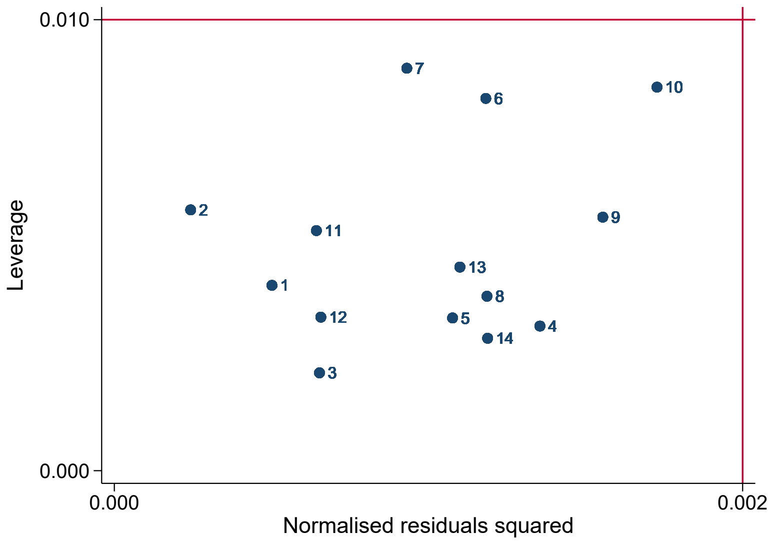
Figure C.17. *Influence analysis in Acemoglu et al.'s (2008)*



NOTE: *Unit 5 is Angola; 54 is East Timor; 56 is Egypt; 105 is Lesotho; 125 is Monzambique; 133 is Nigeria; 135 is Oman; 139 is Panama; 140 is Papal States; 150 is Russia.*

Figure C.18. *Leverage-Residual plot for Schularick and Taylor (2012)*

(a) *Model (2)*



(b) *Model (3)*

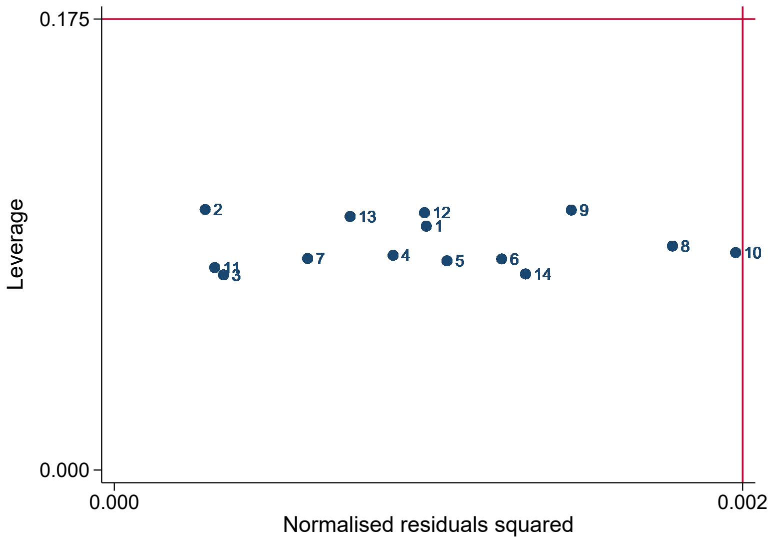
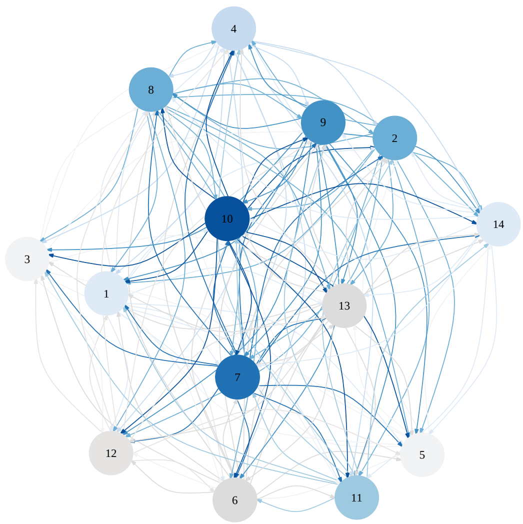


Figure C.19. *Influence analysis in Schularick and Taylor's (2012)*

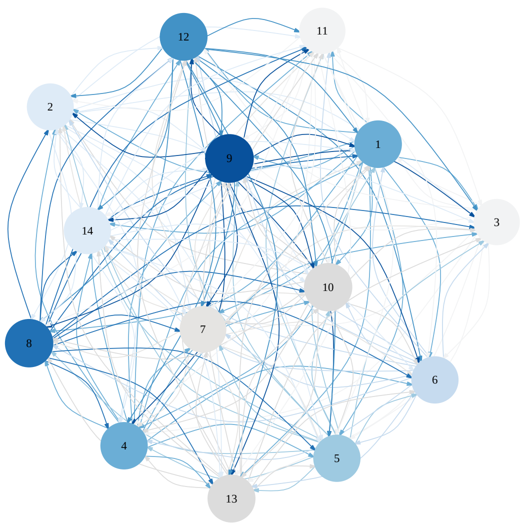
Model (2)



Enhancing effect

Masking effect

Model (3)



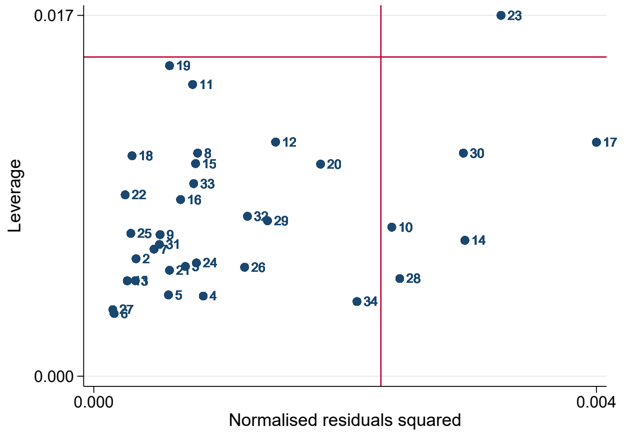
Enhancing effect

Masking effect

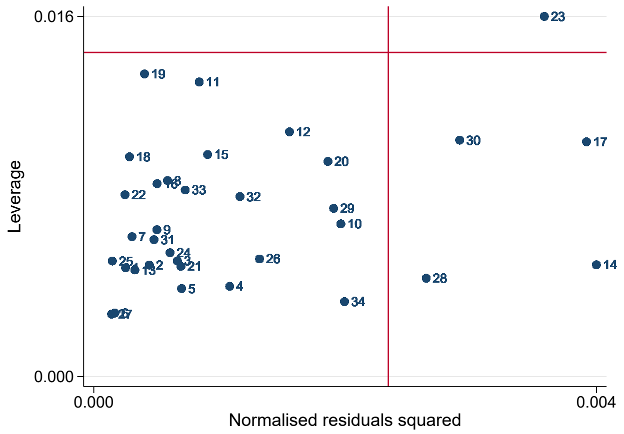
NOTE: *Unit 10 is Japan.*

Figure C.20. *Leverage-Residual plot for Égert (2016)*

(a) *Model (1)*



(b) *Model (2)*



(c) *Model (3)*

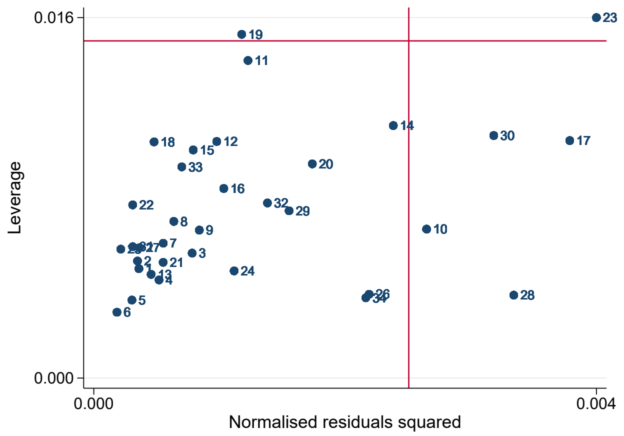
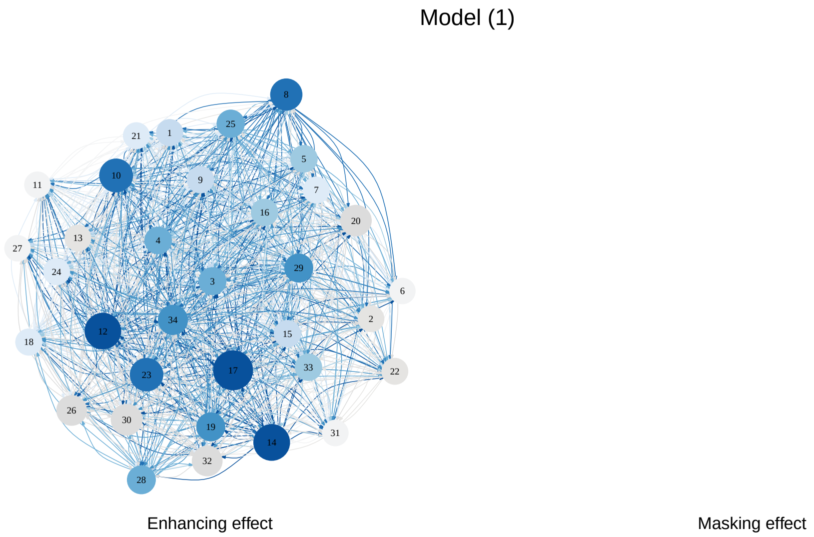
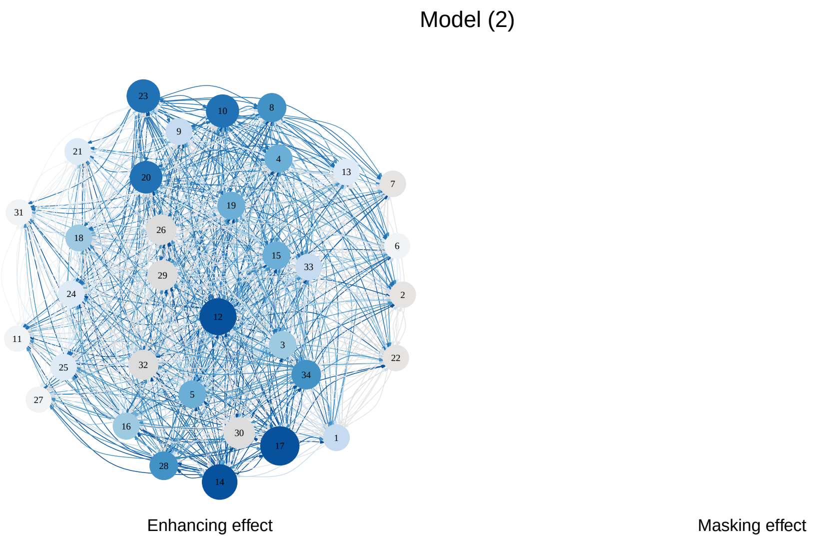
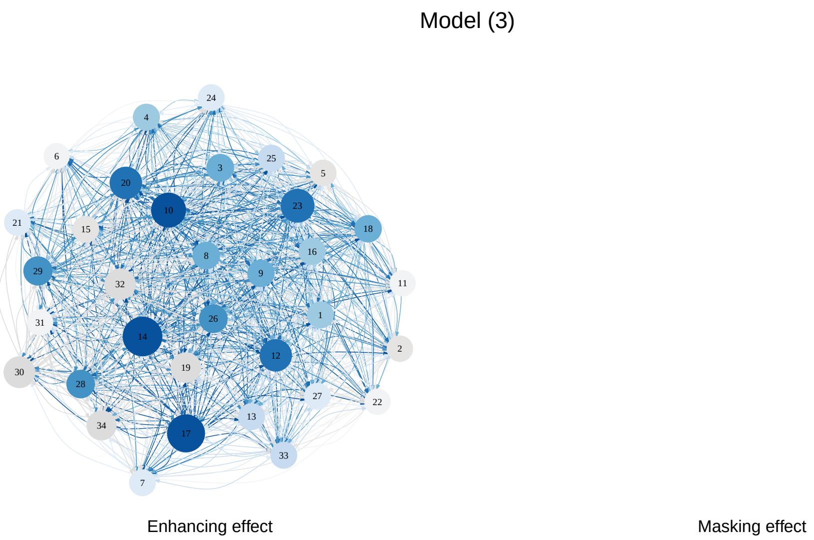


Figure C.21. *Influence analysis in Égert's (2016)*



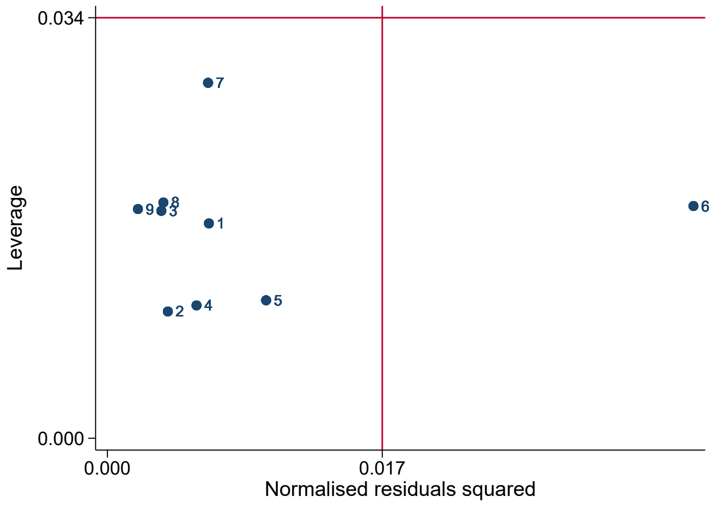




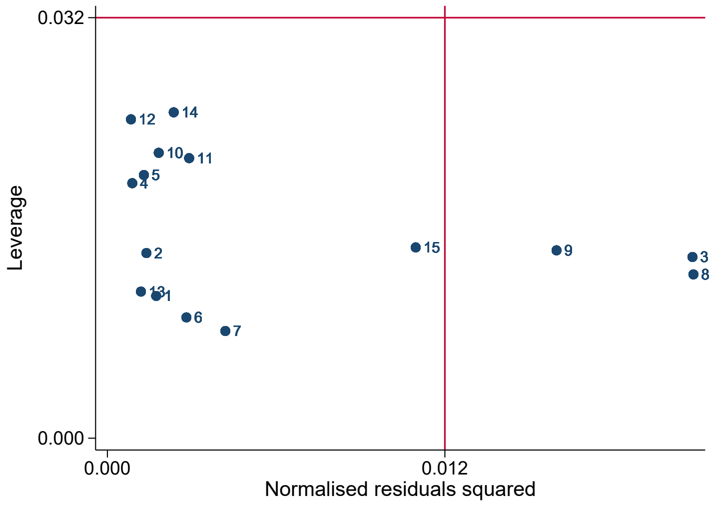
NOTE: *Unit 10 is Spain; unit 14 is Great Britain; unit 17 is Ireland; unit 19 is Israel; unit 23 is Luxemburg; unit 28 is Poland; unit 34 is United States.*

Figure C.22. *Leverage-Residual plot for Berka et al. (2018)*

(a) *Model (2a)*



(b) *Model (2b)*



(c) *Model (2c)*

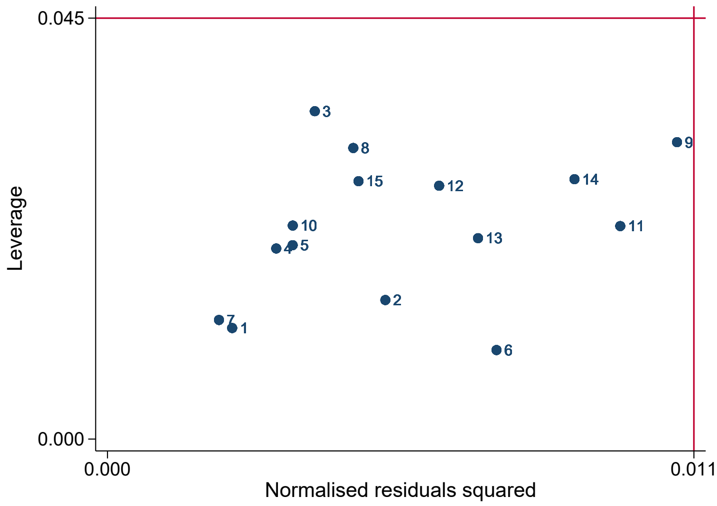
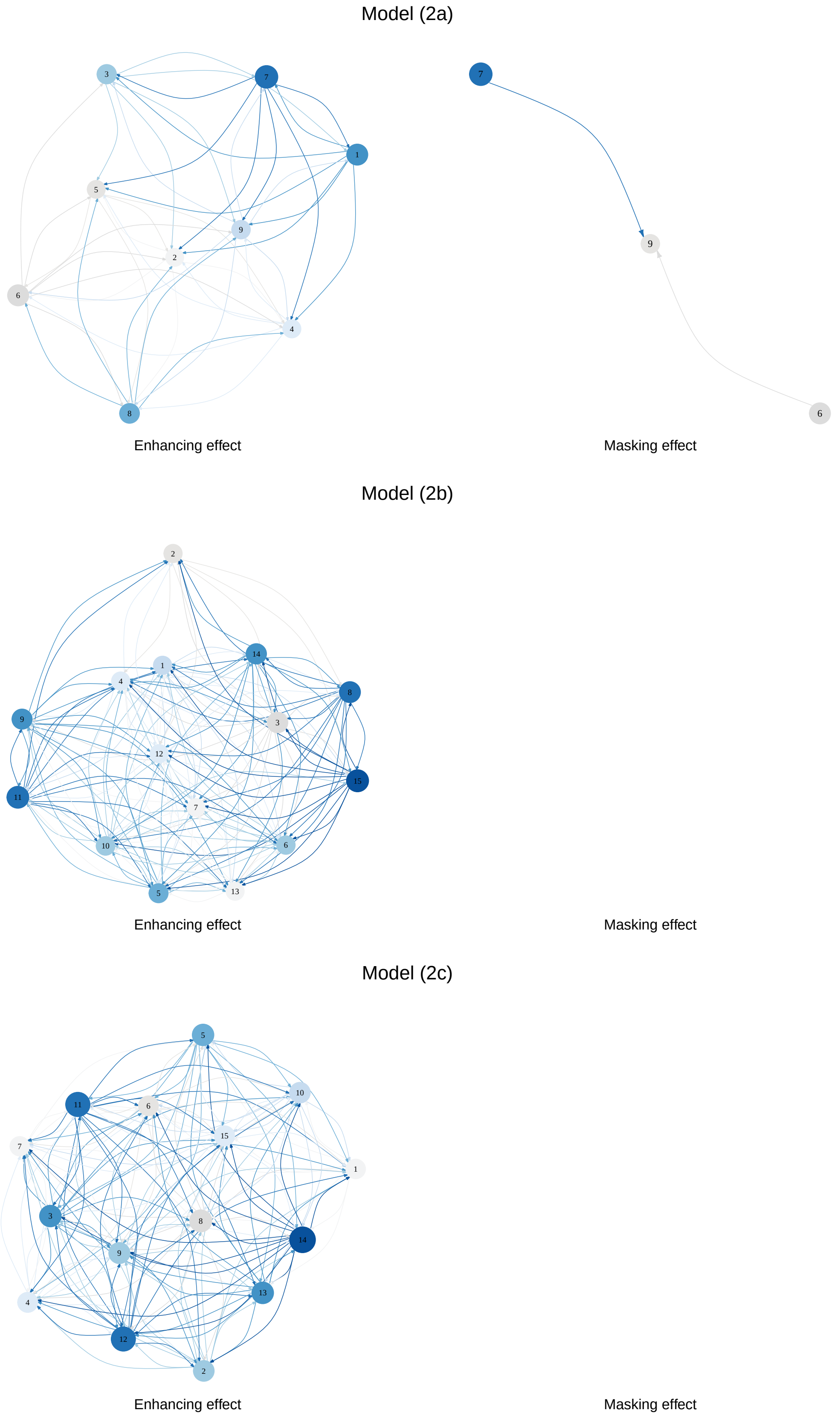


Figure C.23. *Influence analysis in Berka et al.'s (2018)*



NOTE: *Unit 3 is Czech Republic; unit 6 is France; unit 7 is Germany; unit 8 is Hungary; unit 9 is Ireland; unit 15 is United Kingdom.*

Figure C.24. *Labour market outcomes by gender*

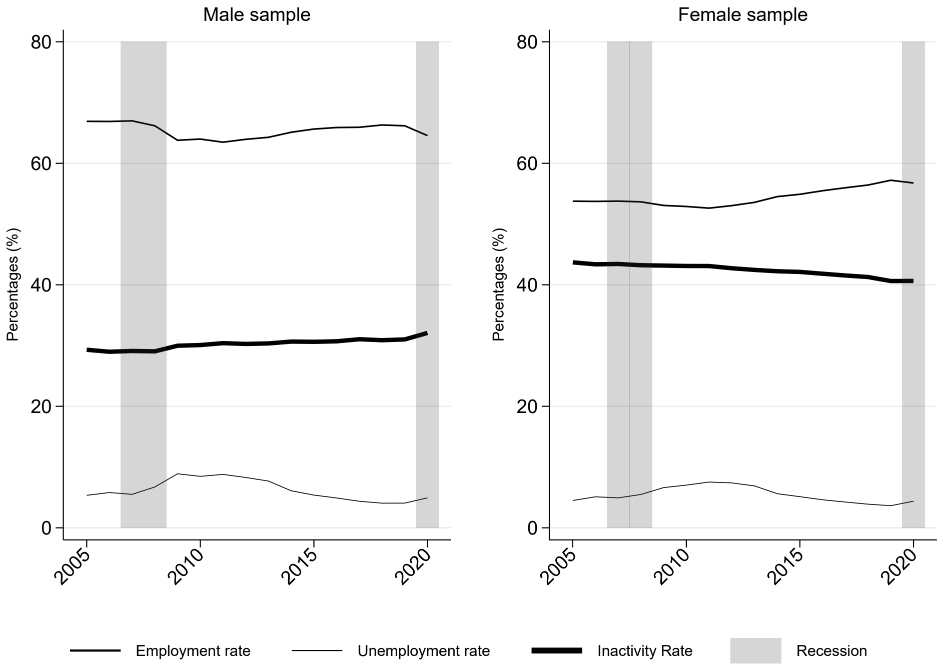


Figure C.25. *Distribution of gender sectoral segregation index, by sectoral dominance*

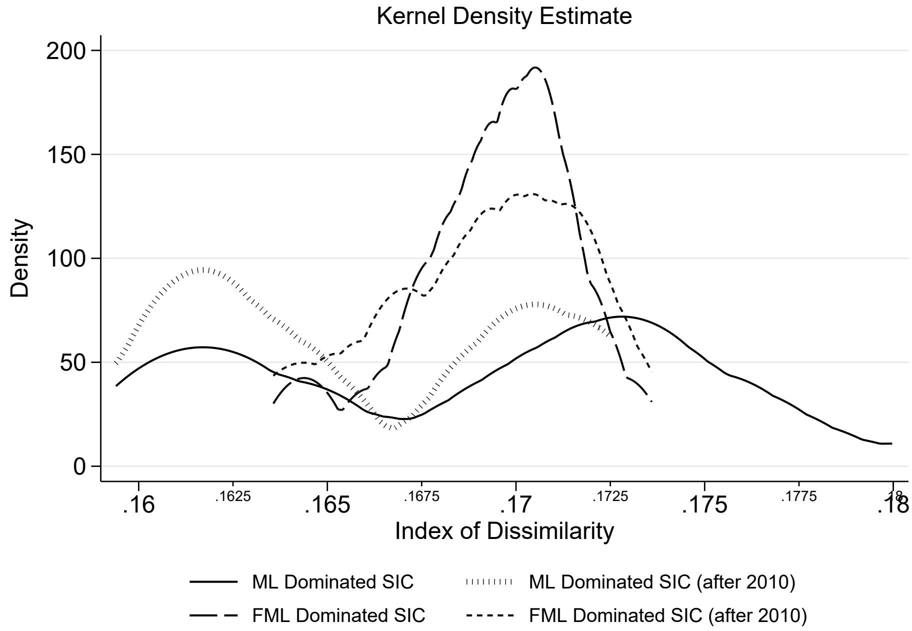


Figure C.26. *Distribution of gender sectoral segregation index, by sectoral dominance and degree of segregation*

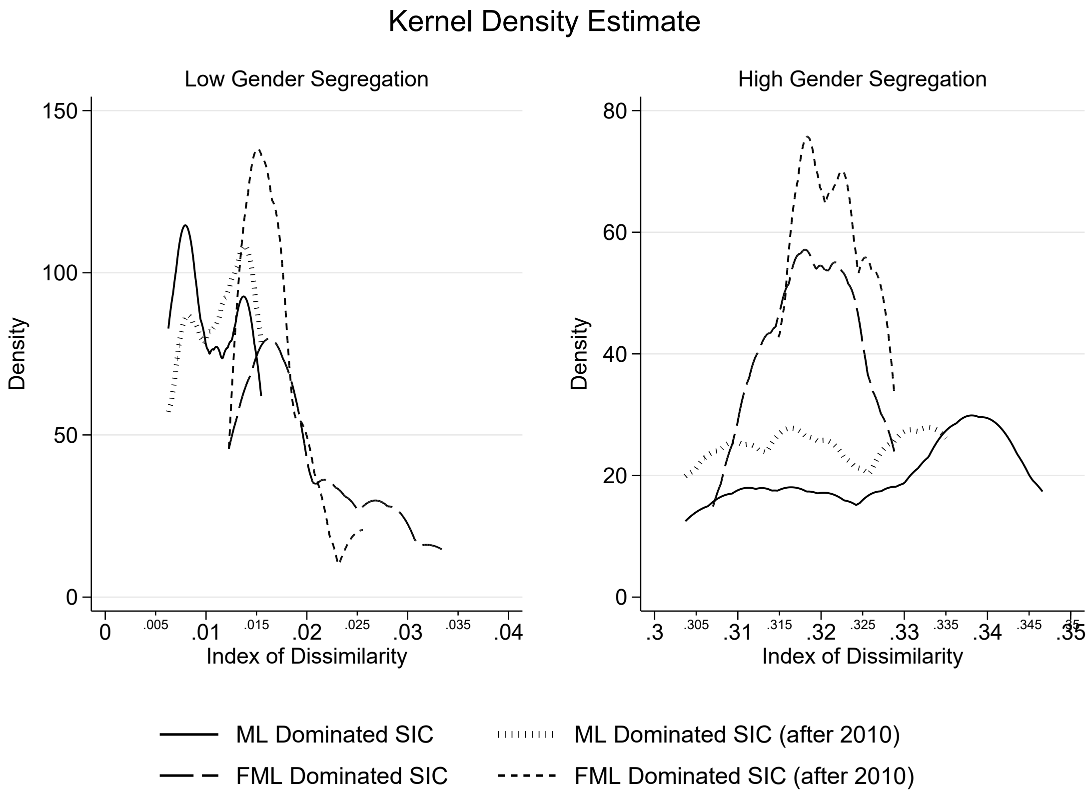
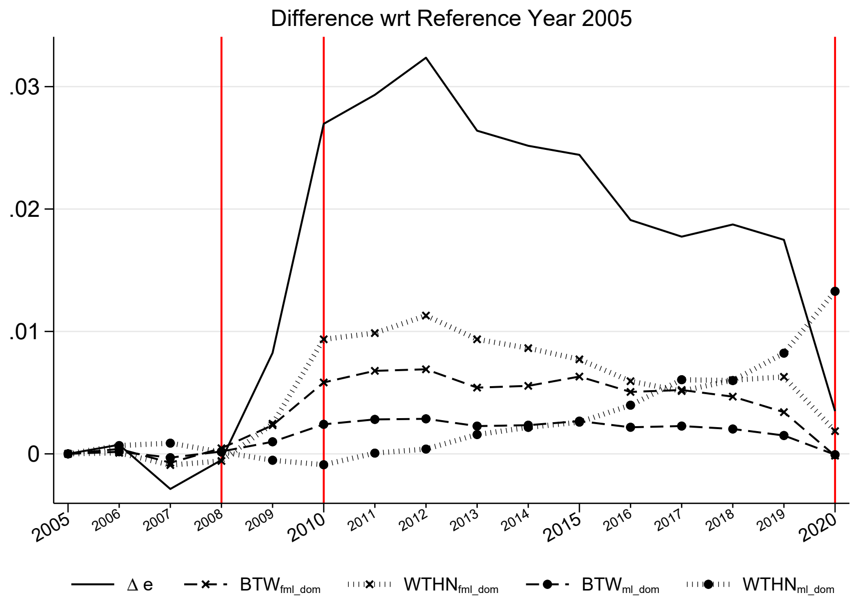
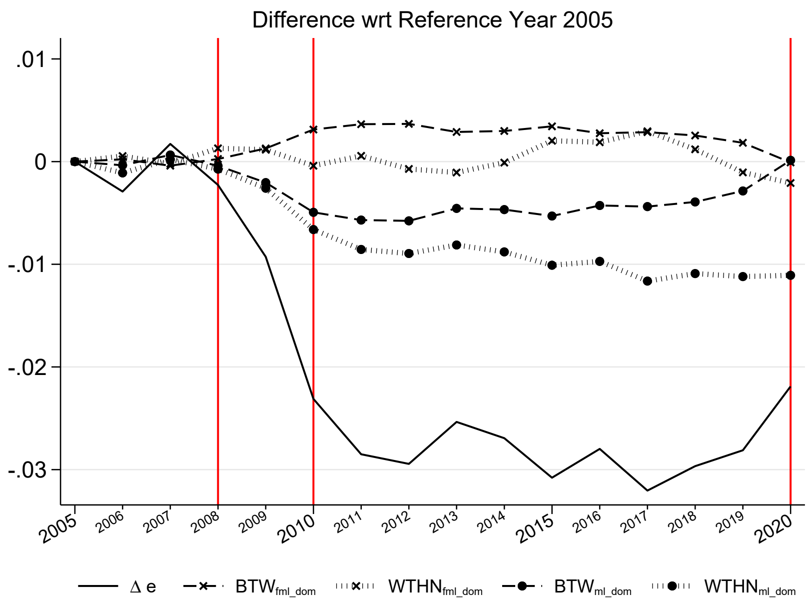


Figure C.27. *Shift-share decomposition of employment, by gender*

(a) *Female sample*

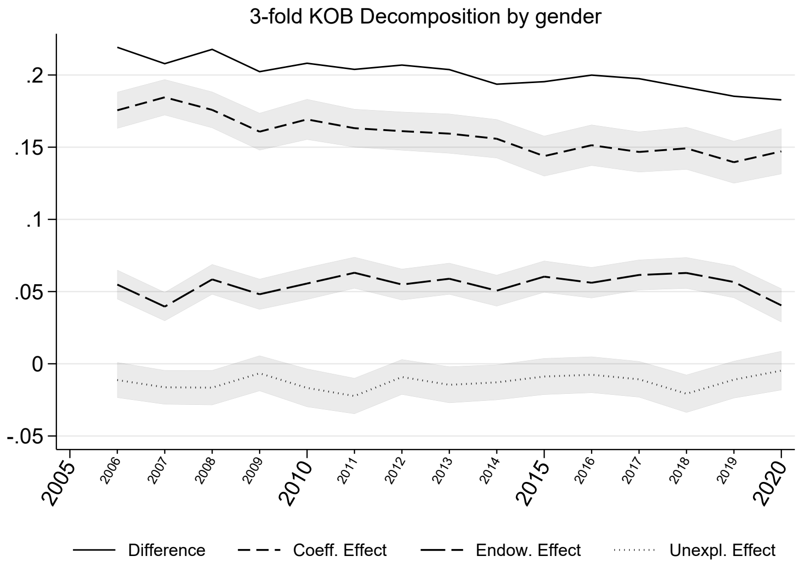


(b) *Male sample*



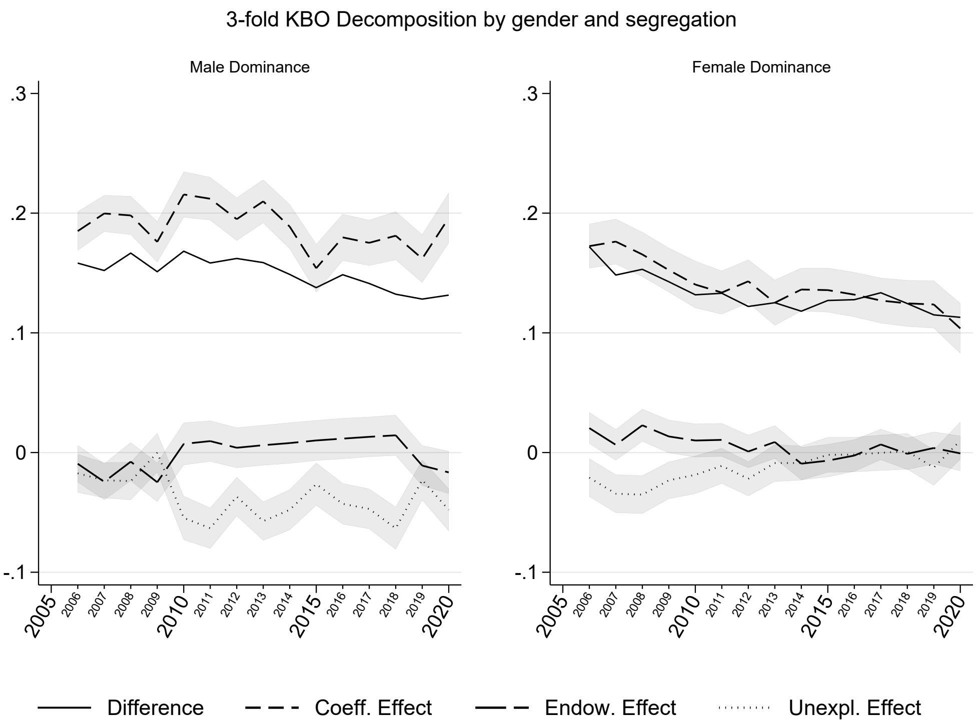
Note: The graphs display the shift-share decomposition of female and male employment (respectively, at the top and bottom). Both graphs show the difference in employment in the comparison year with respect to the base year (i.e., the fiscal year 2005) for women (at the top) and men (at the bottom). The overall change in employment is shown in solid line and its decomposition into the between *and* within *components respectively, with dashed and dotted lines. The cross marks the components for female-dominated sectors and the circle the components for male sectors. The* between *component (BTW) captures the change due to changes in the sectoral structure of the economy; the* within *component (WTHN) reflects changes in female composition within sectors.*

Figure C.28. *3-fold KBO decomposition*



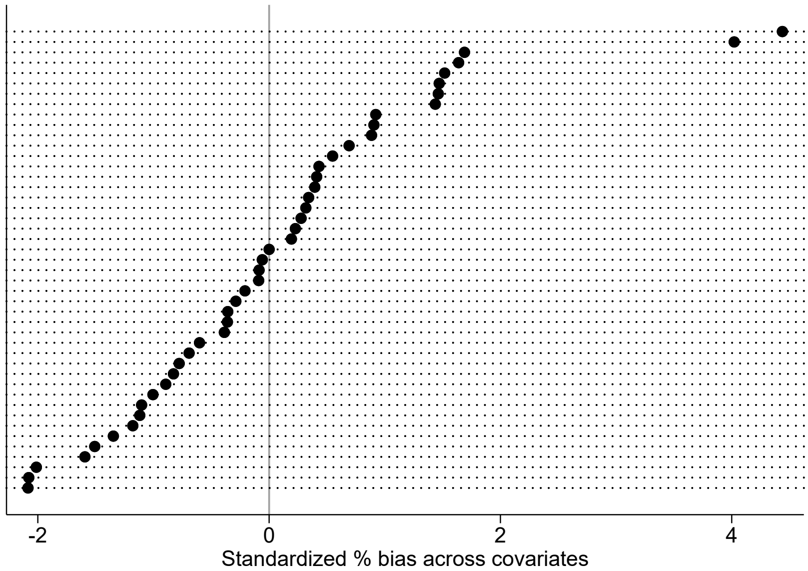
ESTIMATION NOTE*: Both models for women and men are estimated using the Mincerian regression equation (with OLS). The shaded areas is the 95% confidence intervals.*

Figure C.29. *3-fold KBO decomposition, by gender sectoral dominance*



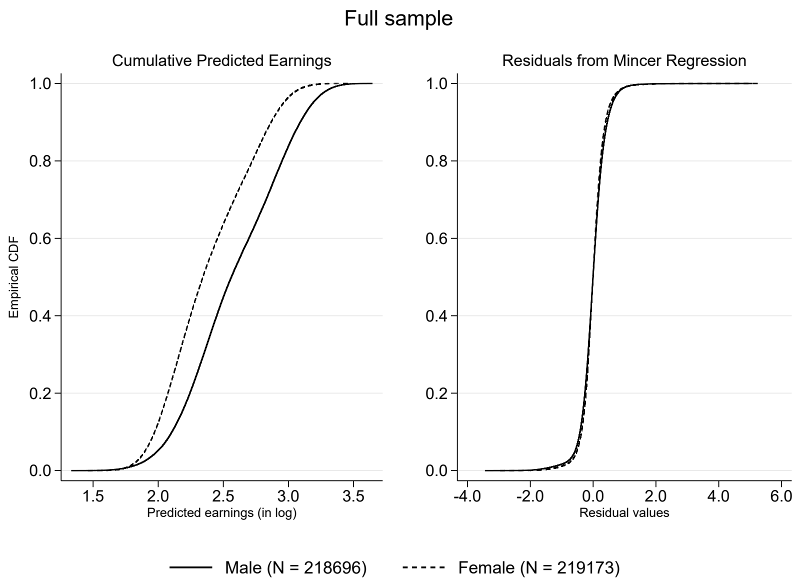
ESTIMATION NOTE*: Both models for women and men are estimated using the Mincerian regression equation (with OLS). The degree of gender segregation is not included because it is highly correlated with the grouping variable of gender sectoral dominance. The shaded areas are the 95% confidence intervals.*

Figure C.30. *Covariate imbalance test, single components*



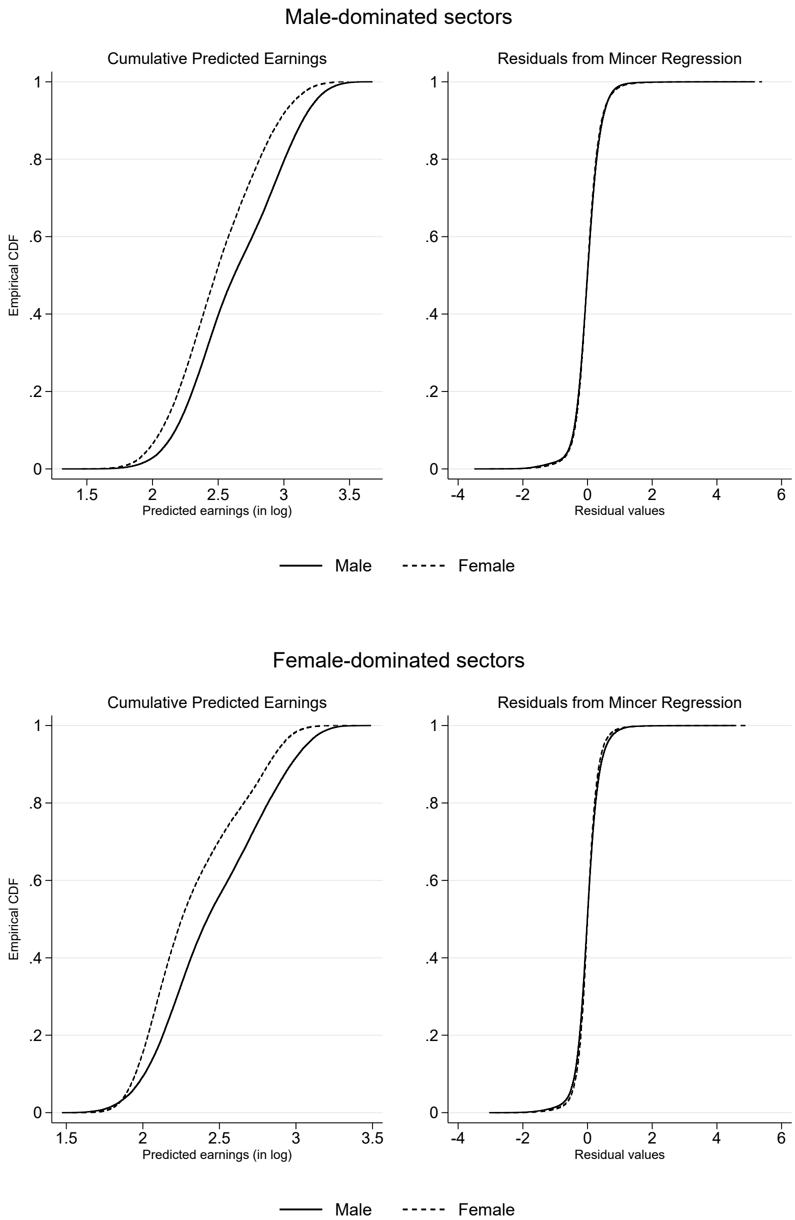
Note: The included covariates are balanced if the standardised bias after matching is within ±5% (Rosenbaum and Rubin, 1985). If the condition is satisfied, the matching method successfully builds a valid control group.

Figure C.31. *CDFs of predicted earnings and residuals, all sectors*



Note: Predicted earnings are a precise measure of individual earnings potential (Gould and Moav, 2016; Borjas et al., 2019), and the residuals from a Mincerian regression to capture the part of earnings that is uncorrelated to observed skills (Parey et al., 2017). Predicted earnings and the residuals from a Mincerian regression are calculated after estimating the coefficients of the Mincerian wage regression, reported in Table B.28.

Figure C.32. *CDFs of predicted earnings and residuals, by gender sectoral dominance*



Note: Predicted earnings are a precise measure of individual earnings potential (Gould and Moav, 2016; Borjas et al., 2019), and the residuals from a Mincerian regression to capture the part of earnings that is uncorrelated to observed skills (Parey et al., 2017). Predicted earnings and the residuals from a Mincerian regression are calculated after estimating the coefficients of the Mincerian wage regression, reported in Table B.28.