UNIVERSITY OF COPENHAGEN
FACULTY OF SCIENCE

BSc in computer science and economics

How a supply and demand shock affects labor market in Denmark

A case study for all faculties at University of Copenhagen

Carl Gustav Frederiksen

Mads Byth Halborg Jensen

Betim Hamidi

Supervised by Jakob Egholt Søgaard

June 2023

Carl Gustav Frederiksen

Mads Byth Halborg Jensen

Betim Hamidi

How a supply and demand shock affects labor market in Denmark

BSc in computer science and economics, June 2023

Supervisors: Jakob Egholt Søgaard

University of Copenhagen

Faculty of Science

Department of Computer Science

Universitetsparken 1

2100 Copenhagen Ø

Preface

This bachelor thesis has been completed as part of a 15 ECTS project at the Department of Computer Science, University of Copenhagen. The decision to explore the topic of how a supply and demand shock affects labor market in Denmark was motivated by the growing importance of understanding the complex dynamics of the labor market and guides policy-making, helps businesses with workforce planning, and informs strategies to enhance economic resilience amid increasing global uncertainties.

The intended reader of this paper is a student with a fundamental understanding of economics and data science, equivalent to a Bachelor's degree. The student should be familiar with basic econometrics concepts such as different estimators, assumptions, demand and supply, economic shocks as well as core principles of labor market analysis.

We would like to express our deepest gratitude to Professor Jakob Egholt Søgaard for their invaluable guidance and supervision throughout this project. Their insightful feedback, encouragement during the idea generation phase, and dedication to ensuring we adhered to the time plan have been instrumental in the successful completion of this thesis. Without their support, this paper could not have come to fruition.

Abstract

Vi undersøger effekterne af udbuds- og efterspørgselschok på arbejdsmarkedet i Danmark for at forstå dynamikken i beskæftigelse og lønfluktuationer. Ved at anvende et omfattende datasæt, fokuserer vi på virkningen af disse chok på beskæftigelse og timelønninger. Studiet afslører, at øget produktion, vores proxy for udbudsschock, øger beskæftigelsen, men sænker timelønningerne. Overraskende nok førte efterspørgselschokket, indikeret ved Tysklands import, til et fald i beskæftigelsen. Virkningen af efterspørgselschokket på timelønningerne kunne dog ikke fastslås på grund af databegrænsninger. Disse resultater understreger de nuancerede virkninger af økonomiske chok på arbejdsmarkedet og opfordrer til yderligere omfattende studier i dette område.

Contents

1 Introduction

The labor market plays a vital role in the economic well-being of a country, and understanding how it responds to external shocks is crucial for policymakers and researchers. This study aims to investigate the effects of supply and demand shocks on the labor market in Denmark, providing valuable insights into the dynamics of employment and wage fluctuations within the Danish economy. Supply and demand shocks can have profound impacts on the overall economic health, labor market dynamics, and standard of living. By studying these relationships, policymakers and businesses can develop more informed strategies to mitigate adverse effects or capitalize on beneficial changes resulting from such shocks.

To achieve our research objectives, we have collected a comprehensive dataset that encompasses a range of relevant variables. The dataset includes information on the number of employed individuals, wage earners, worked hours, hourly wages, production levels, input quantities, production costs, and Germany imports. By utilizing this dataset, we can gain valuable insights into the intricate dynamics between shocks, labor market outcomes, and industry-level performance in Denmark. Our primary focus is to examine the impact of demand and supply shocks on employment and hourly wages. Demand shocks, such as changes in consumer preferences, aggregate demand fluctuations, or shifts in international trade patterns, can significantly influence labor markets. Supply shocks, on the other hand, arise from changes in the availability or cost of inputs, technological advancements, or supply disruptions. These shocks can have profound implications for employment and wage determination within specific industries.

We hypothesize that both demand and supply shocks exert influence on employment and hourly wages, although their impacts may manifest in complex and sometimes counter intuitive ways. By employing rigorous empirical methods, we will identify the causal effects of these shocks while accounting for other relevant factors that may confound the relationship. Our paper reveals complex relationships between demand and supply shocks and labor market outcomes. Specifically, increases in production, our proxy for supply shocks, were found to boost employment but decrease hourly wages. Surprisingly, demand shocks, as indicated by Germany's imports, led to a decrease in employment. However, the influence of demand shocks on hourly wages couldn't be determined due to data limitations. These findings underscore the nuanced impacts of economic shocks on the labor market, calling for further comprehensive studies in this domain.

2 Data overview

We have utilized data compiled from multiple tables sourced from "Danmarks Statistik: Statistikbanken" and "OECD". The data from the Danmarks Statistik: Statistikbanken database covers the period from 1966 to 2019, while the data from the OECD database spans from 1990 to 2019. We have incorporated data from i = 115 distinct industries, each evaluated over a span of T = 53 years, resulting in a total of N = 6210 observations per variable. The data type used in this thesis is referred to as Panel Data. The structure of the data in this study is of significant relevance to our research, as it accommodates both time variant and cross-sectional variations. In the dataset, we've decided to transform the variables into logarithmic form in order to gain data normalization. This transformation also aids in stabilizing the variance across the data spectrum. Furthermore, it compresses the data scale, thereby diminishing the impact of extreme values. In this transformed state, a fixed percentage alteration in the original variable equates to a fixed unit modification. This transformation enhances the interpretability of coefficients. For instance, in a log-log model, the coefficients can be understood as elasticities, representing the percentage change in the dependent variable corresponding to a percentage change in an independent variable.

Dataset exploration

Log employed is a measure of total number of workers within a given industry and year. We include both normal wage earners, who are engaged in paid work or services, whether full-time, part-time, or temporary, in that specific industry and also includes self-employed individuals, volunteers, consultants and other persons who is not working under a salary contract.

Log wage earners are individuals who work for an employer and receive compensation in the form of wages or salaries for the work they perform. This is in contrast to self-employed individuals who work for themselves and receive income based on their business profits or revenues. Wage earners may be employed full-time, part-time, or on a temporary basis.

Log total wages represents the cumulative compensation paid to all employees in a specific industry and year, encompassing salaries, hourly wages, bonuses, commissions, and other benefits.

Log worked hours Denotes the total hours worked by wage earners in a specific industry and year, including regular hours, overtime, and any additional time worked.

Log hourly wage is a linear relationship between Total wages (mil. DKK.) and worked hours (1000 hours). We calculate the hourly wage from $\frac{\textbf{Total wages (mil. DKK)}}{\textbf{worked hours (1000 hours)}}$ and it refers to the amount paid to employees pr. worked hour. The hourly wage is increasing from annually - with a few exceptions, where it is decreasing in some years, but afterwards increasing again.

Log production is a measure of the total value of goods and services produced in a given industry and year. Production is now viewed as a proxy for supply shocks. This is a reasonable assumption because production measures the total value of goods and services produced in a given industry and year, which can be directly affected by supply-side factors such as changes in input costs, technological advancements, or disruptions in the supply chain.

Log Input-output describes the interdependence between different sectors of an economy. The data within is the total value of the export in a given industry and year. Inputs refer to the resources used by a sector to produce its output, such as raw materials, energy, labor, and capital. Outputs refer to the products or services produced by a sector and sold to other sectors or to final consumers. The input-output variable can be used to evaluate the impacts of changes in one sector on the rest of the economy, and mainly we use it to analyze the effect of economic shocks on wages.

Log production cost refers to the total expenses incurred by a company in order to produce and manufacture a product or provide a service. Production costs can include various expenses, such as labor costs, raw material costs, energy costs, facility costs, and equipment costs. We will use this mainly as an explanatory variable, in our thesis.

Log Germany imports signifies the total value of goods imported to Germany from Denmark. Given the significant trade relationship between Germany and Denmark, changes in this variable can reflect demand shocks in Danish industries. However, our data on this variable is somewhat limited, as not all industries have import/export data, and for those that do, the data is not comprehensive. 'Germany Imports' is considered as a proxy for demand shocks.

This is a fair assumption because the total value of goods imported to Germany from Denmark can be seen as a reflection of German demand for Danish goods. Changes in this demand, due to factors such as economic conditions in Germany or changes in consumer preferences, can cause demand shocks in Danish industries.

Descriptive statistics

Table 2.1: Descriptive Statistics

		1970	1977	1984	1991	1998	2005	2012	2019
Wage earners	Mean	17535.5	18479.6	19441.5	20598.3	21747.1	22510.3	22481.8	24559.3
	Std. Dev.	25047.5	2770.6	30663.9	33538	36362.4	37640.9	38351.6	41200.7
Total wages	Mean	567.087	1364.5	2707.55	4042.6	5266.24	6964.64	8500.16	10389.9
	Std. Dev.	866.548	2063.76	4049.78	5952.4	8107.04	10594.9	13181.4	15758.7
Worked hours	Mean	31378.9	28327.9	29124.9	28806.4	30468.6	31797.4	31496.5	33391.2
	Std. Dev.	44614	40801.1	43885.3	44696.9	47795.6	50180.9	50693.8	52790.9
Hourly wage	Mean	19.1288	50.1518	99.7672	150.143	178.927	230.43	282.618	330.614
	Std. Dev.	5.71005	12.0958	23.4267	36.5706	35.8654	52.5361	64.9142	84.082
Production	Mean	1817.29	4205.86	8632.05	12108.8	16006.2	22567.1	28093.6	33946
	Std. Dev.	2606.1	5694.53	10788.2	14037.2	17944.9	26155.8	35071.1	44594.5
Input-output	Mean	270.151	681.577	1715.6	2522.46	3252.62	5397.45	7372.55	9387.7
	Std. Dev.	627.414	1522.58	3529.37	5135.18	6515.43	14048.5	21075.1	27324.9
Production cost	Mean	866.635	2018.58	4215.91	5680.82	7581.22	11339.3	14552.7	17324.9
	Std. Dev.	1429.87	3100.42	5938.22	6981.06	8867.3	14541.8	21548.2	26648.2
Germany import	Mean	0	0	0	213460	200682	309562	390261	344126
	Std. Dev.	0	0	0	603021	570686	866415	1087590	975088

Note: This table presents the mean and standard deviation values of various economic indicators from 1970 to 2019. Notably, there has been a general increase in the mean values of most indicators over the years, with "hourly wage" and "production" exhibiting significant growth. However, it should be noted that there was no data available for "Germany import" from 1966 to 1989, resulting in zero values for the years 1970, 1977, and 1984.

Table 2.1 presented above displays the descriptive statistics for various economic variables, used in our dataset, over the years 1970, 1977, 1984, 1991, 1998, 2005, 2012, and 2019. A mean and standard deviation of values has been provided for each variable during these years. The variables that has been used to construct the table are: "Wage earners", "Total wages", "Worked hours", "Hourly wage", "Production", "Input-output", "Production cost", and "Germany import".

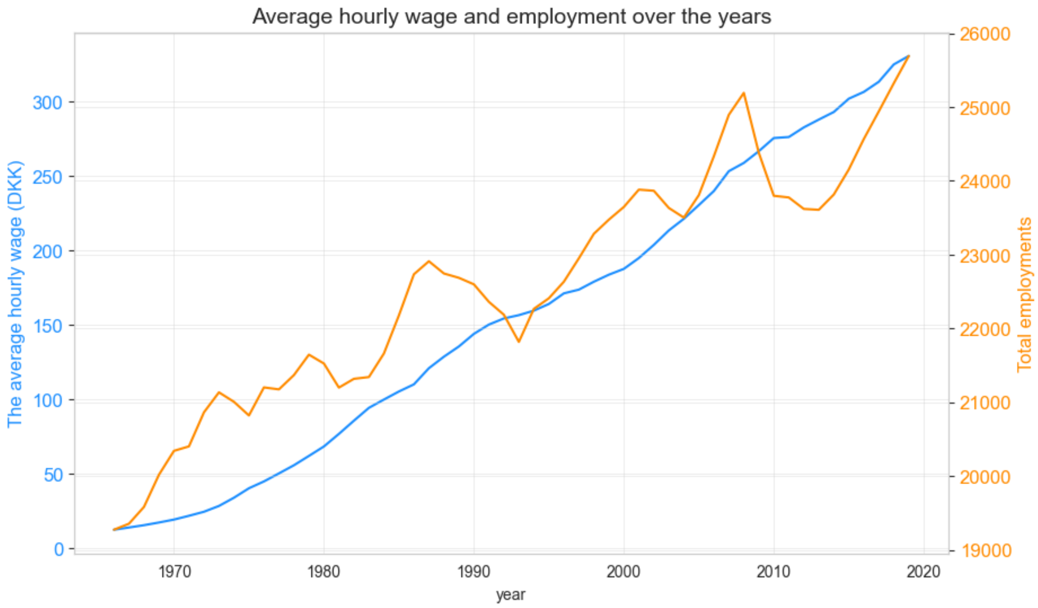
The mean values is a representation of the average values of the variables for the given years, while the standard deviation indicates the spread of the data points around the mean value.

A greater standard deviation indicates a broader spectrum of values and increased variability within the data. A smaller standard deviation suggests a narrower range of values and less variability within the data, signifying that the data points tend to be concentrated near the mean.

Generally, we can tell that there has been an increment for most of the variables over time, which can be seen for the mean value over the years. Other than that, we can tell that there's also an increment for the standard deviation, which is an indication that the variability of the data has grown over the years as well. One notable exception is the "Germany import" variable, which has a 0 mean and standard deviation for the years of 1970, 1977 and 1984. The reason for this is, that there was no data for "Germany import" until the year of 1990.

To summarize the table, it gives a clear overview of the trends and variability of the different economic variables over time, in an interval of 7 years. It highlights the general increase in the mean value and standard deviation values for most variables, that has been used.

Figure 2.1: Development of employment levels and hourly wage throughout the years

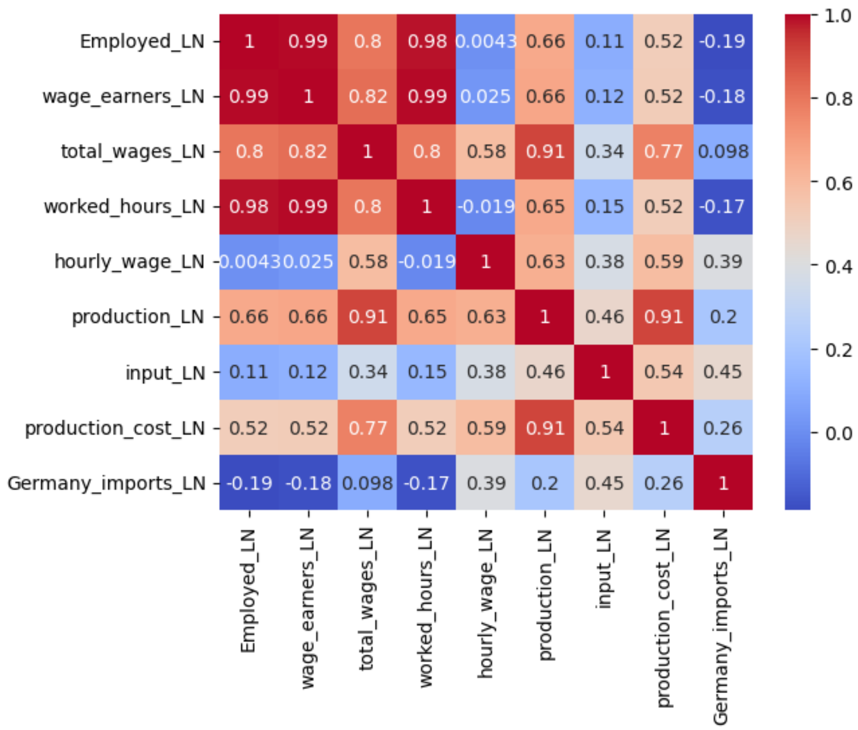


The line plot depict the average hourly wage and total employment in 115 industries in Denmark over the years. The average hourly wage shows a steady increase throughout the entire period, suggesting that wages have been generally improving over time. This could be attributed to various factors, including productivity growth, Germany imports, education and overall wealth.

Total employment exhibits a more dynamic pattern. The decline in employment between

1986-1993 could be due to the economic recession of the early 1990s. The period of employment increasing from 1993 to around 2008 might be related to the economic expansion and globalization that took place. Lastly employment levels falls until 2013-2014, can be attributed to the global financial crisis, which had a significant impact on many economies, including Denmark. These fluctuations in employment levels could be a result of supply or demand shocks affecting different industries in Denmark

Figure 2.2: Correlation Matrix Heatmap



The heatmap displays the correlation matrix between the variables in the dataset, which is providing us with a visualized view of the interrelationship among the different variables. The color scheme to the right represents the strength of the correlation between the variables. It shows, that the cells with a dark blue color gives an indication of weak or negative correlation (under 0), while the cells with a dark red color signify a strong positive correlation between the variables.

We can observe in figure 2.2, that several of our variables gives a strong positive correlation with each other, which is an indication of a close relationship between the variables. It's important to note, that there's strong correlation among the variables 'Employed_LN', 'Wage_earners_LN', 'Total_wages_LN', and 'Production_LN'. When one of these variables experiences an increment, the others, that are correlated with it, tends to increase as well. This reflects their interconnected nature in the economy.

The heatmap highlights the complex connections between different economic factors, with a particular focus on the close ties between labor-related variables and production. The color-coded representation simplifies the process of understanding the relationships among variables, emphasizing the strong positive correlations between wage earners, total wages, worked hours, and production, where all of the individual variables are taken in logarithmic form, as it can help us linearize the relationship.

3 Theoretical framework

We begin by assuming that the industry identifier, which categorizes data by industry, is fixed and stable over time, reflecting the unique characteristics and dynamics of each sector. Moreover, the temporal dimension of our dataset, spanning from 1966 to 2019, is sufficient to capture long-term trends and cyclical fluctuations in the labor market and production processes.

We assume that technological progress affects all industries, but its impact may vary across sectors. For the purposes of our analysis, we will assume that technological progress is constant across all industries. This assumption suggests that the rate of technological change and advancement remains steady over time, regardless of the specific industry in question. This means that all sectors are considered to experience the same pace of technological innovation.

We assume that both labor and capital can move freely between industries, in response to changes in relative wages and returns on capital. This assumption allows for the possibility of labor and capital reallocation in response to shifts in demand and supply.

We assume that the industries in our dataset operate under conditions of perfect competition. This implies that there are a large number of buyers and sellers in each industry, products are homogeneous, firms have perfect information about prices and production techniques, and there are no barriers to entry or exit. This assumption simplifies the analysis and allows us to focus on the direct effects of demand and supply shocks on wages and employment.

We assume that trade between Denmark and Germany is unrestricted, with no tariffs, quotas, or other barriers to the movement of goods and services across borders. This assumption is consistent with the fact that both countries are members of the European Union, which promotes free trade among its member states. Free trade facilitates the transmission of demand and supply shocks between the two countries and affects the wages and employment in Denmark's industries.

Dynamics

In this section, we outline the assumptions for each variable in our dataset, which will help us better understand their relationships and dynamics within the theoretical framework.

We assume that the total wages paid to all employees reflect the industry's labor productivity and are influenced by factors such as bargaining power, competition, and the regulatory environment. We would expect a positive relationship with hourly wage, as industries that pay higher total wages may also pay higher hourly wages. For employment, the relationship could be positive if higher total wages attract more workers to the industry, or negative if high wages are a result of a labor shortage.

Meanwhile, the worked hours variable we would assume employees in an industry work more hours, this could indicate higher labor demand and potentially higher wages. However, it could also indicate that employees are working longer hours due to low wages, which would suggest a negative relationship. If employees in an industry work more hours, this could indicate higher labor demand and potentially higher wages. However, if employees are working longer hours due to low wages, it could suggest a negative relationship between worked hours and hourly wage. For employment, the relationship could be positive if more hours worked indicates more workers, or negative if more hours worked indicates overtime by existing employees.

Turning to production-related variables, we assume that the production value is a measure of an industry's output and economic contribution, determined by factors such as technology, capital investment, and labor input. Production we assume can be seen proxy for a supply shock.

The production cost variable refers to the total expenses incurred by a company in order to produce and manufacture a product or provide a service. In a regression analysis, we assume a positive relationship between production cost and hourly wage, as industries with higher production costs might need to pay higher wages to attract and retain skilled workers. However, if companies are trying to cut costs due to high production expenses, they might limit wage growth, suggesting a negative relationship. For employment, the relationship could be negative if high production costs lead companies to cut back on hiring or even lay off workers. Conversely, industries with high production costs might also be those with high production levels, which could require more workers, suggesting a positive relationship.

The input-output variable describes the interdependence between different sectors of an economy. It could have complex relationships with employment and wages, depending on the balance of inputs and outputs in the industry. For example, an industry that is more dependent on inputs from other sectors may be more vulnerable to supply shocks, which could affect employment and wages. If an industry relies heavily on inputs from other sectors and there's a disruption in the supply of those inputs, it might need to cut back on production, which could lead to layoffs and lower wages. On the other hand, if an industry produces a lot of output that is used by other sectors, a surge in demand from those sectors could lead to increased production, higher employment, and potentially higher wages.

Finally, we assume that the Germany imports, which measures the total value of goods imported to Germany from Denmark for each industry, is driven by factors such as trade policies, exchange rates, and the overall economic relationship between the two countries it is seen as our proxy for a demand shock. Additionally, we assume that the free trade agreement between Germany and Denmark as members of the European Union allows for relatively unrestricted movement of goods and services, which could influence the volume of imports between the two nations.

We could also acknowledge the possibility of non-linear relationships between variables. For instance, the relationship between wages and employment could be non-linear, with wages potentially rising at a decreasing or increasing rate as employment increases. Likewise, the impact of supply and demand shocks on wages and employment may also be non-linear, depending on the size of the shock and other economic factors. We want to simplify our model, therefore, we won't allow for the non-linear relationships.

Demand and supply

In this section, we outline the demand and supply assumptions that will be incorporated into our model estimation framework. These assumptions ensure that our analysis of the effects of demand and supply shocks on wages and employment at the industry level is consistent and theoretically sound.

We assume that the demand and supply curves for goods and services produced by each industry in our dataset are linear functions of price, with constant slopes representing the price elasticity of demand and supply, respectively, we can represent relationships as:

$$Demand : Q_d = a - bP$$

$$Supply : Q_s = c + dP$$

where Q_d and Q_s represent the quantity demanded and supplied, P is the price, and a, b, c, and d are constants. The equilibrium condition is given by the equality of quantity demanded and quantity supplied:

$$Q_d = Q_s$$

This leads to the equilibrium price:

$$P* = (a - c)/(b + d)$$

Given the equilibrium price, we can determine the equilibrium quantity, which is the production level:

$$Q* = a - bP*$$

We initially assumed perfect competition across all industries, however, we recognize that some industries may exhibit characteristics of oligopoly or monopolistic competition. In these situations, firms have some degree of market power, and prices are determined not just by supply and demand but also by strategic considerations. This could influence the transmission and impact of supply and demand shocks on wages and employment. Therefore, we will attempt to account for market structure in our model, possibly by including industry-specific factors or allowing for different effects across industries.

A positive supply shock (e.g., an increase in production efficiency) will result in a decrease in the equilibrium price and an increase in the equilibrium quantity. This can be represented by a downward shift in the supply curve, with a new constant c':

$$Q'_s = c' + dP$$

where $c' < c$. The new equilibrium price and quantity are:

$$P*' = (a - c')/(b + d)$$

$$Q*' = a - bP*'$$

In this scenario, we would expect the increased production to lead to an increase in employment, as more labor is needed to produce the higher output. However, the effect on wages is ambiguous, as the lower price might put downward pressure on wages, while the increased demand for labor might lead to higher wages.

A positive demand shock (e.g., an increase in demand from Germany) will result in an increase in both the equilibrium price and quantity. This can be represented by an upward shift in the demand curve, with a new constant a': Demand after shock:

$$Q'_d = a' - bP$$

where $a' > a$. The new equilibrium price and quantity are:

$$P*' = (a' - c)/(b + d)$$

$$Q*' = a' - bP*'$$

In this scenario, we would expect both employment and wages to increase, as the higher price and output lead to a higher demand for labor and a willingness of firms to pay higher wages.

Economic shock

Economic shocks refer to unexpected events or changes in the economic environment that affect the supply or demand for goods and services in our dataset. These shocks could originate from various sources, such as changes in government policies, technological innovations, or external events that impact trade, production, or labor markets.

There are two types of economic shocks: demand shocks and supply shocks. Demand shocks occur when there are sudden changes in the levels of demand for goods and services, while supply shocks happen when there's an unexpected change in the ability to produce goods and services. Both types of shocks can either be positive (increases) or negative (decreases).

A positive demand shock would unexpected increase in demand, such as a surge in German imports from Denmark, would typically result in higher prices and higher quantity produced and sold. With increased production, firms would likely need more labor, leading to increased employment. Wages could also increase due to the higher demand for labor. Meanwhile a negative demand shock conversely, would unexpected decrease in demand could result in lower prices and lower production levels. This could lead to a decrease in employment as firms scale back production. Wages may also decrease as the demand for labor falls.

Furthmore a positive supply shock in the production could be due to an increase in the efficiency of production, a decrease in input costs, or a technological advancement that allows the industry to produce more with the same resources. In this scenario, we would expect output and employment to increase as more goods and services are produced and more workers are needed to meet this increased production. The effect on wages is more ambiguous: they could increase due to the higher demand for labor, but they could also decrease if the increase in output results in an oversupply of goods, leading to lower prices and potentially lower wages. Lastly, a negative supply shock this could be due to a decrease in the efficiency of production, an increase in input costs, or a disruption in the supply chain that hampers the industry's ability to produce goods and services. In this scenario, we would expect output and employment to decrease as firms scale back production due to the increased costs or supply disruptions. Wages could also decrease due to the reduced demand for labor, but they could also increase if the reduction in output leads to a shortage of goods, resulting in higher prices and potentially higher wages.

4

Econometric Method

In this section, our research focuses on introducing the methodology we employ. We aim to determine the appropriate estimator and the assumptions it must satisfy for us to draw causal conclusions. Specifically, we will consider two models: Pooled Ordinary Least Squares and Fixed Effects models, which align with our dataset characteristics. As our data exhibits penal data characteristics and correlation between unobserved factors and both the dependent variable and the explanatory variables, we lean towards fixed effects models rather than random effects models. Additionally, we will explore various techniques for model comparison and selection, including the F-test, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). By employing these techniques, we aim to provide a comprehensive understanding of statistical modeling in linear regression analysis.

Pooled OLS

The pooled OLS (ordinary least squares) estimator is a method for estimating the parameters of a linear regression model when there is no reason to suspect that the relationships between the variables differ across individuals or groups. In other words, it assumes that the same linear relationship holds for all individuals or groups in the sample.

The model can be expressed as follows:

$$y_{it} = \beta_0 + \beta_1 X_{it1} + \beta_2 X_{it2} + ... + \beta_k X_{itk} + u_{it} \quad t = 1, 2, \cdots, t \tag{4.1}$$

where y_{it} is the outcome variable for individual i in time period t, X_{it} is a vector of explanatory variables, β is a vector of coefficients to be estimated, and u_{it} is the error term.

The pooled OLS estimator estimates the coefficients β using all of the available data, regardless of the individual or group to which each observation belongs. Specifically, it minimizes the sum of squared residuals across all individuals and time periods:

$$min_{\beta_0,\beta_1,\cdots,\beta_k} = \sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \beta_0 - \beta_1 X_{it1} - \beta_2 X_{it2} - \cdots - \beta_k X_{itk})^2 \tag{4.2}$$

We make assumption for the model and estimator to be unbiased and consistent, such that we can estimate causal effects.

Assumptions

MLR.1 is a linearity assumption. The relationship between the dependent and independent variable is linear, whereas the model captures this linear relationship by estimating the coefficient of the independent variables. The relationship is represented by the following equation:

$$y_{it} = \beta_0 + \beta_1 X_{it1} + \cdots + \beta_k X_{itk} + u_{it} \tag{4.3}$$

where β values are the unknown parameters, and u is the unobserved random error term. The linearity assumption provide flexible model, as the underlying variables of interest can be arbitrary functions of the original data, such as logarithmic, square etc.

MLR.2 is an assumption of random sampling in the model. We assume that the data in our model represent a random sample of the population, which can be used to estimate the parameters. In essence, this means that we believe the data we are using in our model are not cherry-picked or influenced by any sort of selection bias, but rather, they constitute a random sample drawn from the larger population of interest. It allows us to employ probabilistic theory to make statistical inferences about the parameters we are trying to estimate. In addition it gives us confidence that the relationships we observe in our data sample are representative of those in the larger population, and therefore, the conclusions we draw from our analysis can be generalized to that population.

MLR.3 assumes no perfect collinearity between the independent variables. This means that no independent variable is a constant value. In other words, there is some variation in every explanatory variable. This makes sense because if one of these variables remained the same for all observations, it wouldn't be very useful in explaining changes in the dependent variable. Moreover, there isn't an exact linear relationship between any two or more independent variables. This means that you can't precisely predict the value of one explanatory variable using a linear equation of the others. If this was possible, it would suggest that these variables are conveying the same information, which could lead to problems in identifying the individual impact of each variable on the dependent variable.

MLR.4 is the assumption of zero conditional mean. The error term, u, has an expected value of zero given any value in the explanatory variables.

$$E(u_{it}|X_{it1}, X_{it2}, \cdots, X_{itk}) = 0$$

this assumption ensures that our unobserved factors are, on average, unrelated to the explanatory variables. This is key in deriving the first statistical property of each OLS estimator: unbiasedness. Under assumptions MLR.1 through MLR.4 the POLS is unbiasedness and consistent, meaning we can reach causal effect in the output.

MLR.5 is an assumption for homoskedasticity. This assumption states, that the error term has a constant variance across all observations. If it happens, that there may be a violation of this assumption, and there is homoskedasticity, then the estimates will still be unbiased and consistent, but the standard error of the estimates will be incorrect. This may affect the hypothesis testing and confidence intervals. The relationship is represented by the following equation, that the error term has a constant variance:

$$Var(u_{it}|X_{it1}, X_{it2}, \cdots, X_{itk}) = \sigma^2$$

MLR.6 is a assumption of normality in the error term, u. We assume that the error term is independent of the explanatory variables, and that it is normally distributed. this assumption imply a stronger efficiency property, but can be discarded is we have a large sample size. under MLR.6 the OLS estimators have the smallest variance among all unbiased estimators.

Even though all assumptions are important, it will not necessary lead to unreliable estimates, if some of them are violated. For the POLS-model, there's clearly some assumptions, that are more important than others. The most important assumptions, for the model to show unbiased and consistent estimates, are MLR.1 through MLR.4. While MLR.5 and MLR.6 are important assumptions as well, they may not lead to the estimates to be biased, if they are violated.

FE-model

Fixed effects estimation is a method for estimating the parameters of a linear regression model when there are unobserved time-invariant characteristics that vary across units (e.g., individuals, firms, countries). These unobserved characteristics are often referred to as "fixed effects," and they can cause problems for traditional OLS regression analysis because they violate the assumption of independence of observations.

The fixed effects estimator accounts for these unobserved characteristics by including a separate parameter for each unit in the model. This parameter, we'll denote as a_i, represents the unit-specific intercept or "fixed effect." The model is then estimated using within-unit variation, which effectively removes the unit-specific effects from the data.

More formally, consider a linear regression model with N units and T time periods:

$$y_{it} = \beta X_{it} + a_i + u_{it} \tag{4.4}$$

where y_{it} is the outcome variable for unit i at time t, X_{it} is a vector of explanatory variables, β is a vector of coefficients to be estimated, and u_{it} is the error term. The fixed effect α_i represents the unobserved time-invariant characteristics of unit i.

The fixed effects estimator estimates the coefficients β by taking the differences between each unit's observations and their own mean:

$$\hat{\beta}_{FE} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it} - \bar{X}_i)(y_{it} - \bar{y}_i)\right)\left(\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it} - \bar{X}_i)^2\right)^{-1} \tag{4.5}$$

where $\bar{X}i$ and $\bar{y}i$ are the sample means of Xit and yit, respectively, for unit i.

This estimator is consistent, unbiased, and efficient under certain assumptions.

Assumptions

FE.1 is a linearity assumption. We assume that for each i=industries the model is:

$$y_{it} = \beta_1 X_{it1} + \beta_2 X_{it2} + \cdots + \beta_k X_{itk} + a_i + u_{it}, t = 1, 2, \cdots, t \tag{4.6}$$

Where β_j are the parameters we estimate, and a_i is the unobserved effect. This model assumes that the dependent variable y_{it} is a linear function of a set of independent variables, $X_{it1}, X_{it2}, ..., X_{itk}$, plus an unobserved or latent variable, a_i, that is specific to each industry i. The unobserved effect, a_i, is also known as the individual fixed effect, and it captures all the time-invariant characteristics of the industries that are not directly included in the model

FE.2 is a data assumption. We assume that we have a random sample from all cross sections. Specifically, it assumes that the data is obtained from a random sample of all possible cross sections or groups in the population of interest.

FE.3 is a explanatory variable assumption. We assume that the explanatory variable changes over time (at least for some i), and we assume that there is no perfect linear relationships among the explanatory variables exists. The assumption that some explanatory variables vary over time is important because it allows the fixed effects model to capture the within-group variation in the dependent variable that is not explained by the time-invariant individual fixed effects.

FE.4 is an error term assumption. We assume that for each t=time periods, the expected value of the idiosyncratic error given the explanatory variables in all time periods and the unobserved effect is zero:

$$E(u_{it}|X_i.a_i) = 0$$

Under these first four assumptions, the fixed effect estimator is unbiased. The key is strict exogeneity assumption, FE.4. Under these assumptions the fixed effect estimator is consistent with a fixed T as N$\rightarrow \infty$

FE.5 is a homoskedasticity assumption. We assume that the amount of variation in the error term u_{it} that is not explained by the observed variables X_i and the time-invariant heterogeneity a_i is the same for all time periods for each individual.

$$Var(u_{it}|X_i, a_i) = Var(u_{it}) = \sigma^2$$

FE.6 is a error terms assumption, regarding serially correlation. We assume that for all t $\neq$ s, the idiosyncratic error are uncorrelated(conditional on all explanatory variables and a_i):

$$Cov(u_{it}, u_{is}|X_i, a_i) = 0$$

Under assumption FE.1 through FE.6 the fixed effects estimator of the β_j is the best linear unbiased estimator(BLUE). The key assumption that makes the Fixed effect estimator the best, is FE.6 which implies that the idiosyncratic errors are serially uncorrelated.

FE.7 is a error term assumption. Conditional on X_i and a_i, the u_{it} are independent and identically distributed as Normal(0,σ_u^2). Assumption FE.7 implies FE.4, FE.5 and FE.6 but it is stronger because it assume Normal distribution for the idiosyncratic errors. Without FE.7 we can rely on asymptotic approximations, that require large N=number of industries and small t=time periods values

Model testing and information criteria

The F-test for variable exclusion, is a statistical test used to assess whether a subset of variables in a regression model can be excluded without significantly reducing the model's fit.

The F-test for variable exclusion compares the fit of a full model, which includes all the variables of interest, with the fit of a reduced model, which excludes one or more variables. The test evaluates whether the exclusion of those variables leads to a significant decrease in the overall model fit.

The F-test is based on the ratio of two sums of squared residuals: the sum of squared residuals of the full model($SSR_{complex}$) and for the reduced model(SSR_{simple}). The test statistic is calculated as:

$$F = \frac{(SSR_{simple} - SSR_{complex})/q}{SSR_{complex}/(n - k_{complex-1})}$$

where $q = k_{complex} - k_{simple}$ and n is the number of observations.

The resulting F-statistic is compared to the critical value from the F-distribution with degrees of freedom (q, n - k - 1) to determine the statistical significance. If the F-statistic exceeds the critical value, it indicates that excluding the subset of variables significantly reduces the model fit, suggesting that those variables are important for explaining the variation in the dependent variable.

The Akaike Information Criterion (AIC) is a statistical measure used to assess the quality of a regression model and compare different models. Its purpose is to find the model that strikes the best balance between goodness of fit and model complexity. AIC is calculated using the following formula:

$$AIC = -2 * log - likelihood + 2 * k$$

In this equation, the log-likelihood represents a measure of how well the model fits the data, and k denotes the number of parameters in the model.

The AIC value is derived by taking negative two times the log-likelihood and adding two times the number of parameters. The log-likelihood captures the likelihood of observing the given data under the assumptions of the model. By penalizing models with more parameters through the inclusion of the 2k term, AIC addresses the trade-off between model complexity and goodness of fit. The model with the lowest AIC value is generally preferred as it suggests a better fit to the data while considering model complexity.

The Bayesian Information Criterion (BIC), is another statistical measure used to assess the quality of a regression model and compare different models. Similar to AIC, BIC aims to find the model that strikes a balance between goodness of fit and model complexity. BIC is calculated using the following formula:

$$BIC = -2 * log - likelihood + k * log(n) * k$$

In this equation, the log-likelihood represents the measure of how well the model fits the data, k denotes the number of parameters in the model, and n represents the sample size.

The BIC value is derived by taking negative two times the log-likelihood and adding the natural logarithm of the sample size multiplied by the number of parameters. BIC penalizes models with more parameters more heavily compared to AIC due to the inclusion of the ln(n) term, where n represents the sample size.

Similar to AIC, a lower BIC value indicates a better-fitting model. However, compared to AIC, BIC tends to penalize complex models more strongly, which can result in a more parsimonious model selection. BIC puts a greater emphasis on model simplicity, encouraging the selection of models with fewer parameters.

5 Results

FE - Employment

Table 5.1

Dependent variable: Log of Employed	
Model	(6) FE
	Coefficient (p-value) [std.err.]
Log of production	0.0404 (0.0000) [0.0061]
Log of Germany import	-0.0040 (0.0000) [0.0003]
Log of Production cost	0.0400 (0.0000) [0.0050]
Log of Input	-0.0079 (0.0000) [0.0009]
Log of Total wages	-
Log of Worked hours	0.8805 (0.0000) [0.0034]
R^2	0.9770
F-statistics	5.138e+04
F-test	1.11e-16
Log-likelihood	7518.5
AIC	-15026.99
BIC	-14993.32

Table 5.1 is a summary output of our fixed effects model with employment as the dependent variable.

The log of production coefficient is 0.9254. This means that a 1% increase in production is associated with a 0.9254% increase in employment, all else being equal. The standard error for this coefficient is 0.0176, which is relatively small, indicating a high degree of precision in the estimate of this coefficient. The p-value is 0.0000, which is less than the common significance level (0.05), suggesting that the coefficient for the log of production is statistically significantly different from zero. Therefore, production appears to have a significant positive impact on employment.

The log of Germany import coefficient is -0.0272, indicating that a 1% increase in Germany import is associated with a 0.0272% decrease in employment, all else being equal. The standard error is 0.0011, and the p-value is 0.0000. Thus, the coefficient is statistically significant, and the negative sign suggests that increased import from Germany has a small but significant negative impact on employment.

The log of production cost coefficient is -0.0947. A 1% increase in production cost is associated with a 0.0947% decrease in employment, all else being equal. The standard error of this coefficient is 0.0173, and the p-value is 0.0000, indicating that the effect of production cost on employment is statistically significant.

The log of input coefficient is -0.0079, indicating that a 1% increase in input is associated with a 0.0079% decrease in employment, all else being equal. The standard error is 0.0009, and the p-value is 0.0000, indicating that the effect of input on employment is statistically significant.

The log of worked hours coefficient is 0.8805, indicating that a 1% increase in worked hours is associated with a 0.8805% increase in employment, all else being equal. The standard error is 0.0034, and the p-value is 0.0000, which is less than the common significance level (0.05). Therefore, the effect of worked hours on employment is statistically significant.

The p-values for all variables are low, which is typically interpreted to mean that there is very strong evidence against the null hypothesis that the coefficient is zero (i.e., that the variable has no effect on the dependent variable). A common threshold for "statistical significance" is $p < 0.05$, and all variables here are well below this threshold.

Correlation with model dynamics

Generally the model aligns with some of the expectations set forth in the dynamics. The positive coefficient in the regression model indicates that increases in production costs correspond to increases in employment. This suggests that firms might respond to rising production costs by hiring more labor, perhaps to maintain or increase productivity. Another possibility is that labor is being substituted for other, more expensive, inputs to production. The exact mechanism would likely depend on specific industry dynamics, including the availability and relative costs of different inputs, the potential for automation, and the nature of the product being produced. The negative coefficient for the 'Log of Input' variable could imply several dynamics at play. It may suggest that companies are becoming more efficient, possibly through technological improvements or better organizational practices, thus requiring less labor to produce the same level of output. Alternatively, it might indicate that firms are substituting labor with other inputs, possibly due to changes in relative input prices or due to strategic choices to optimize production.

The log of Germany imports variable has a negative coefficient. This is contradictory to the expectation that an increase in demand from Germany (higher imports) would increase employment in the Danish industry. A negative coefficient suggests that an increase in Germany's demand for goods is associated with a decrease in employment in the industry under study. The negative coefficient in the model, can seem counterintuitive at first glance. However, there are several potential explanations for this outcome: If firms in Denmark are responding to increased demand from Germany by automating their production processes or utilizing more technology, then this could potentially lead to less demand for labor, hence the decrease in employment. Another explanation could be companies may decide to move production or certain stages of production to other countries where labor or other production costs are cheaper. This can happen if an increase in demand pushes companies to seek ways to produce more efficiently. Lastly, another solution could be that, danish firms could be importing more intermediate goods from Germany for final assembly or production in Denmark. This could mean that the increased demand is satisfied through the use of imported materials rather than domestic labor, leading to a decrease in employment.

Production is viewed as a supply shock in this context, the positive coefficient for the 'Log of Production' suggests that an upturn in production requires more labor, resulting in increased employment. This outcome is in line with the theory that an industry in expansion, due to, say, breakthroughs in technology, access to new markets, or favorable policies, will require more labor.

Lastly worked hours, this variable is positively correlated with employment. This is expected because as employment rises, the total number of hours worked in the industry would also tend to increase. However, this relationship could be complicated by factors such as part-time work, overtime, and the use of flexible working arrangements.

The panel data regression model with both entity and time effects aligns with the theoretical framework which assumes that industries have unique characteristics over time and that the model will capture these effects.

The table presents the estimates from a Fixed Effects (FE) model, a type of regression model that is especially useful when dealing with panel data, that is, data that involves observations over time on the same subjects. The dependent variable here is 'Log of Employed', indicating that the model is estimating the logarithm of the number of employed individuals. This transformation can help in dealing with skewed data and often allows for a more linear relationship between predictors and the outcome, which is one of the assumptions of regression models.

While the model seems to account for a substantial proportion of the variation in employment (as indicated by the R-squared value of 0.9970), there could still be other factors not included in the model that influence employment levels. These omitted variables could introduce bias into the estimates if they are correlated with both the dependent variable and any of the independent variables in the model. This is a potential issue because the omitted variables may be causing the observed changes in employment rather than, or in addition to, the predictors currently included in the model.

The relationships assumed between the predictors and the dependent variable is another area of potential concern. The model currently assumes a linear relationship in logs (i.e., a log-log model) between the predictors and employment. If the actual relationships are nonlinear, the model's estimations and inferences could be inaccurate.

Finally, even though the R-squared value is high and the F-test shows overall significance, it's critical to understand that these are not definitive proof of the model's appropriateness or correctness. The R-squared only shows the proportion of the variation in the dependent variable that is explained by the independent variables in the model. It doesn't say anything about whether the right variables have been included, whether the relationships are correctly specified, or whether the assumptions of the model are met. Similarly, the F-test only tests the null hypothesis that all the coefficients are equal to zero against the alternative that at least one is not zero. It doesn't provide information about individual coefficients or potential issues like multicollinearity.

Thus, while these results provide a promising starting point, further investigation is indeed necessary to confirm that the model is well-specified, the assumptions are met, and potential issues like omitted variables and incorrect functional forms are properly addressed.

FE - Hourly wage

Table 5.2

Dependent variable: Log of Hourly Wage	
Model	(5) FE
	Coefficient (p-value) [std. err.]
Log of Production	-0.0169 (0.0137) [0.0069]
Log of Wage earners	-0.7431 (0.0000) [0.0093]
Log of Worked hours	-
Log of Germany import	-
Log of Production cost	0.0168 (0.0027) [0.0056]
Log of Total wages	0.7344 (0.0000) [0.0097]
Log of Input	-
R^2	0.5196
F-statistics	1632.6
F-test	1.11e-16
Log-likelihood	6732.0
AIC	-13456.03
BIC	-13429.10

The table shows the results of a model with hourly wage as explanatory variable and five independent variables. The effects of these independent variables on the dependent variable are measured by coefficients.

The production coefficient of -0.0169 suggests that if the log of production increases by one unit, the log of the hourly wage is expected to decrease by 0.0169 units, holding all other variables constant. Given the log-log model, this translates into an approximate -0.0169% change in the hourly wage for a 1% increase in production. The standard error of 0.0069 is an estimate of the standard deviation of the coefficient. It reflects the level of uncertainty around the estimate of the coefficient. Smaller standard errors suggest that the sample estimate is likely to be close to the population parameter. The p-value of 0.0137 is less than 0.05, which typically signifies statistical significance at the 5% level. This means we would reject the null hypothesis that this coefficient equals zero, suggesting that production has a significant impact on the hourly wage.

The wage earners coefficient of -0.7431 indicates that a one-unit increase in the log of wage earners is associated with a decrease of 0.7431 in the log of the hourly wage, all else being equal. This translates to approximately a -0.7431% decrease in the hourly wage for a 1%.The standard error is 0.0093. A small standard error relative to the coefficient suggests that the estimate of the coefficient is precise and likely to be close to the true population value. The p-value is 0.0000, which is below any common significance level. This means that the number of wage earners has a statistically significant effect on the hourly wage.

The production cost coefficient of 0.0168 suggests that a one-unit increase in the log of production cost is associated with an increase of 0.0168 in the log of the hourly wage, all else being equal. This equates to approximately a 0.0168% increase in the hourly wage for a 1% increase in the production cost. The standard error is 0.0056. The relatively small standard error suggests that the estimate of the coefficient is precise. The p-value is 0.0027, which is less than 0.05, indicating statistical significance

The total wages the coefficient of 0.7344 suggests that a one-unit increase in the log of total wages is associated with an increase of 0.7344 in the log of the hourly wage, all else being equal. This translates to approximately a 0.7344% increase in the hourly wage for a 1% increase in total wages. The standard error is 0.0097. A small standard error indicates that the coefficient is precisely estimated, providing confidence in the estimate. The p-value is 0.0000, indicating that total wages have a statistically significant effect on the hourly wage at any commonly used significance level.

Correlation with model dynamics

We assumed that an increase in total wages would correspondingly increase employment, which could be interpreted as either an increase in labor demand or a decrease in labor supply. In our model, the positive coefficient on total wages aligns with this assumption, suggesting that as total wages increase, the average hourly wage also increases. This could mean that as firms pay more in total wages (either due to hiring more workers or paying higher wages), the average wage per worker also increases.

The assumption was not explicitly stated, but one could assume that if the number of wage earners increases, this could signify a higher demand for labor and therefore potentially an increase in wages. However, our model suggests the opposite: as the number of wage earners (supply of labor) increases, the average hourly wage decreases, which is in line with economic theory that suggests an increase in supply (with demand held constant) leads to a decrease in price, in this case, wages.

In theory, an increase in production, i.e., a positive supply shock, can lead to an increase in employment as companies may require more labor to meet increased production. However, how this affects wages can be more nuanced. If the production increase is due to technological advancements or efficiency gains, firms may not need to increase wages significantly as they can produce more with the same labor force. Conversely, if the increased production is due to an increase in demand for the product, it could lead to higher wages as firms try to attract more workers to increase production.

The log of production has a negative coefficient, indicating that an increase in production is associated with a decrease in the average hourly wage, holding all other variables constant. This outcome contradicts the usual assumption. One potential explanation for this outcome could be that as production increases, firms may be investing more in technology or equipment to increase output, reducing the need for additional labor and thereby putting downward pressure on wages. Another explanation could be the presence of economies of scale, where firms become more efficient as they produce more, and this increased efficiency could lead to a decrease in average wages.

The impact of production cost on wages depends on the nature of these costs. If the increase in production cost is driven by an increase in wages, then it's natural to see a positive relationship between production cost and wages. However, if the production cost is driven by factors unrelated to wages (such as the cost of raw materials), then the relationship isn't as straightforward.

The log of production cost has a positive coefficient, suggesting that an increase in production costs is associated with an increase in the average hourly wage, all else being equal. This could be interpreted to mean that higher production costs are leading to higher wages. If higher production costs are due to higher labor costs, this result would be expected. Alternatively, if production costs are rising due to reasons other than wages (like increased raw material costs), firms may need to offer higher wages to retain employees amid higher operational costs. These workers might need to take on additional responsibilities or skills to handle the new production methods or circumstances leading to the increased costs.

Table 5.2 presents the estimates from a Fixed Effects (FE) model, where the dependent variable used is 'Log of Hourly wage'. The transformation, which takes the variables in logarithmic form helps in dealing with skewed data, which allows in a more linear relationship between the predictors and the outcome.

First of all, we can see that the independent variable 'Log of Wage earners' is negative and also highly significant. We would maybe expect 'Log of Wage earners' to contribute more to 'Log of Hourly wage', because of increased demand for labor, as an example. The estimate for 'Log of Wage earners' may therefore be seen as counter-intuitive.

Secondly, the significance levels of the coefficients need to be taken into consideration. The 'Log of Production' variable has a p-value of 0.0137, which may be considered significant depending on the threshold set for statistical significance (commonly $p < 0.05$). However, in the realm of economic research, it is often the case that we seek more stringent levels of significance, such as $p < 0.01$, in which case, this variable would not be deemed significantly different from zero.

Also, if we look at the R-squared value, we can see that the R-squared value is at 0.5196, which indicates that it explains 51.96% of the variance in the 'Log of Hourly Wage'. It might be considered, that the R-squared value is quite low, and therefore it may be seen as an insufficient fit.

POLS - Employment

Table 5.3

Dependent variable: Log of Employed	
Model	(5) POLS
	Coefficient (p-value) [std.err.]
Log of production	1.0296 (0.0000) [0.0265]
Log of Production cost	-0.3253 (0.0000) [0.0190]
Log of Germany import	-0.0661 (0.0000) [0.0029]
Log of Input	-0.0579 (0.0000) [0.0050]
Log of Total wages	0.4399 (0.0000) [0.0203]
Log of Worked hours	-
R^2	0.9993
F-statistics	1.901e+06
F-test	1.11e-16
Log-likelihood	100.00
AIC	-190.00
BIC	-156.33

The table provides coefficients, p-values, and standard errors for each independent variable, among other things.

The log of production coefficient is 1.0296. This suggests that a 1% increase in production is associated with approximately a 1.0296% increase in employment level. The p-value of 0.0000 suggests this effect is statistically significant, and the small standard error of 0.0265 suggests that this estimate is precise. Essentially, this indicates that increasing production is a viable strategy to increase employment.

The log of Production Cost coefficient is -0.3253. This suggests that a 1% increase in production cost is associated with a 0.3253% decrease in employment. This suggests that as the cost of production increases, employment decreases, assuming all other variables are held constant. The p-value of 0.0000 indicates that this effect is statistically significant, and the standard error of 0.0190 shows that the estimate is precise.

The log of Germany Import coefficient is -0.0661. This suggests that a 1% increase in Germany's import is associated with a 0.0661% decrease in employment. The p-value of 0.0000 shows this effect is statistically significant, and the standard error of 0.0029 shows a high precision of this estimate.

The log of Input coefficient is -0.0579, indicating that a 1% increase in input is associated with a 0.0579% decrease in employment. The p-value of 0.0000 indicates that this effect is statistically significant, and the standard error of 0.0050 shows a high precision of this estimate.

The log of total wages coefficient is 0.4399. This suggests that a 1% increase in total wages is associated with a 0.4399% increase in employment level. The p-value of 0.0000 shows that this effect is statistically significant, and the standard error of 0.0203 suggests that this estimate is precise.

Correlation with model dynamics

The model's results generally align with the assumptions. The negative coefficient (-0.0579) of the log of input suggests that an increase in input is associated with a decrease in employment. This could be interpreted as an increase in efficiency or productivity, where more is being produced with the same or less amount of labor. Alternatively, it could reflect a shift towards more capital-intensive production methods.

The positive coefficient for the log of total wages supports the assumption that an increase in total wages reflects the industry's labor productivity and potentially increases the average hourly wage. This suggests that industries that are capable of paying higher total wages also pay higher hourly wages, which could be due to higher labor productivity or other factors such as more bargaining power for workers or higher demand for labor, and log of production variables support the assumptions that an increase in wages and production corresponds to an increase in employment, assuming all other factors are held constant. However, the negative coefficient for the log of production cost contradicts the initial assumption that an increase in production costs might lead to higher employment levels. It seems that as production costs rise, firms might try to cut down on costs, possibly by reducing the number of employees.

The link between the log of Germany imports and the log of production variables are particularly interesting when examining the coefficients. The negative coefficient for the log of Germany imports contradicts the expected positive impact of a demand shock on employment. This could be due to firms in Denmark responding to increased demand by automating their production processes, or potentially moving parts of their production abroad where production costs might be lower. This could result in a decrease in employment in Danish industries.

The positive coefficient of the log of production suggests that, as production increases, the hourly wage rate also increases. This matches with the general assumption that industries with more output should pay higher wages, possibly due to the increased demand for labor. However, it might seem counterintuitive since in theory, a positive supply shock could lower prices, potentially putting downward pressure on wages. However, if we interpret a rise in production as an improvement in labor productivity, this could indeed lead to higher wages, as productivity is a key determinant of wages in the long run

The coefficients for the log of production and the log of Germany imports variables provide insights into the effects of supply and demand shocks, respectively, on employment. An interesting point to note is that the effects of these shocks appear to be in opposite directions; supply shocks (increased production) appear to increase employment, while demand shocks (increased imports from Germany) appear to decrease employment.

There's different potential issues, that could occur for this table and its estimates. When looking at the model, we have a very high R-squared value, which is estimated at 0.9993, which indicates that the 99.93% of the variance in the dependent variable 'Log of Employed' can be explained with the used independent variables. The R-squared value indicates, that it might fit our model too well, and that the model was made specifically for this data. Therefore, when used for new data, it may perform poorly, because of how specific it is.

Another potential issue, that may occur could be that the variable of 'Log of Production' and 'Log of Production cost' are too closely related to each other. It's entirely plausible that these factors are intricately linked - for example, as production increases, it's likely that production costs would also rise, given the increased expenditure on resources, labor, and possibly even technology or infrastructure. This interrelation, in statistical terms, is known as multicollinearity. Multicollinearity is when two or more independent variables in a regression model are highly correlated. It's important to note that while correlation among variables is almost inevitable in any real-world model, a high degree of correlation (i.e., multicollinearity) can pose challenges.

If multicollinearity is present, it can inflate the variances of the parameter estimates and make them highly sensitive to minor changes in the model. This can, in turn, result in unreliable and unstable estimates of the regression coefficients. The instability of these coefficients could lead to incorrect interpretations and unreliable predictions. In other words, if 'Production' and 'Production cost' are too closely related, a small change in how we measure or define these variables could cause a major shift in the results, which can make it challenging to draw firm conclusions.

POLS - Hourly wage

Table 5.4

Dependent variable: Log of Hourly wages	
Model	(6) POLS
	Coefficient (p-value) [std.err.]
Log of production	0.9635 (0.0000) [0.0161]
Log of Production cost	-0.1968 (0.0240) [0.0129]
Log of Germany import	0.0371 (0.0000) [0.0020]
Log of Input	-0.0245 (0.0000) [0.0033]
Log of Total wages	-
Log of Worked hours	-0.2243 (0.0000) [0.0078]
R^2	0.9786
F-statistics	5.667e+04
F-test	1.11e-16
Log-likelihood	-6603.2
AIC	13216.47
BIC	13250.13

The table presented here is a regression output table from a PooledOLS model where the dependent variable is the log of hourly wages.

The coefficient of Log of production is 0.9635, which implies that if the log of production increases by 1%, the log of hourly wages will increase by about 0.9635%. This is under the condition that all other variables are kept constant. The standard error associated with this coefficient is quite small (0.0161), which means that the estimated coefficient is relatively precise. Also, the p-value for this variable is 0.0000, which is less than the common significance level of 0.05. This means that we reject the null hypothesis that this coefficient is zero, and it is statistically significant.

Next, The coefficient of Log of Production cost is -0.1968. This negative value indicates an negative relationship with the log of hourly wages. Specifically, if the log of production cost increases by 1%, the log of hourly wages is expected to decrease by about 0.1968%, assuming other variables remain constant. Its standard error is 0.0129, which is quite small, indicating a relatively precise estimate. The p-value is 0.0240, which is less than 0.05, hence the effect of this variable is considered statistically significant.

Turning to the coefficient of Log of Germany import is 0.0371, suggesting a direct relationship with the log of hourly wages. In this case, if Germany's import (log-transformed) increases by 1%, we expect the log of hourly wages to increase by about 0.0371%, keeping everything else constant. The standard error associated with this coefficient is extremely small (0.0020), implying a high level of precision in this estimate. The p-value for this variable is 0.0000, which is less than the usual significance level of 0.05. Thus, we can reject the null hypothesis and claim that this coefficient is statistically significant.

The coefficient of Log of Input is -0.0245. This suggests that if the log of input rises by 1%, the log of hourly wages would decrease by about 0.0245%, assuming all other variables are kept constant. Its standard error is 0.0033, indicating a relatively precise estimate. The p-value is 0.0000, less than 0.05, hence the effect of this variable is statistically significant.

The coefficient of Log of Worked hours is -0.2243, indicating an negative relationship with the log of hourly wages. If the log of worked hours increases by 1%, the log of hourly wages would decrease by 0.2243%, with all other variables being constant. The standard error associated with this coefficient is 0.0078, which is relatively small, implying a high level of precision in the estimate. The p-value is 0.0000, less than 0.05, so we can reject the null hypothesis of coefficients being zero and say this coefficient is statistically significant.

Correlation with model dynamics

The production is seen as a proxy for supply shocks, this variable returned a significant and positive coefficient. This suggests that an increase in production is linked to an increase in hourly wages. The nearly one-to-one relationship indicates a highly elastic response of wages to changes in production. This aligns well with the assumptions, where we posited that an increase in production might lead to a higher demand for labor, thus pushing wages up. The real-world situation could be that industries ramp up production and thus require more labor. To attract the necessary workforce, they may offer higher wages. The production cost has a negative coefficient indicates that when the cost of production increases, hourly wages decrease. This aligns somewhat with the assumption where production costs are influenced by labor costs. In practice, when production costs rise, firms may try to offset these costs. Since labor cost is a major component of overall production costs, firms might decrease wages, hence the negative relationship observed. However, it's also worth noting that in the long term, persistently low wages could lead to decreased morale and productivity or increased employee turnover.

Germany imports is seen as a proxy for demand shocks, the variable is positive in this case so this indicates that an increase in imports from Germany is associated with an increase in hourly wages in Denmark. In the context of the assumptions, this result is coherent. We assumed that a positive demand shock, such as an increase in demand from Germany, would result in an increase in both the equilibrium price and quantity. Given that more goods are being sold (higher quantity), it is also expected that more labor is required to meet this demand. As a result, firms might be willing to pay higher wages to attract or retain the labor necessary for increased production. Also, the higher price resulting from increased demand might generate more revenue for firms. This could allow them to afford higher wages. Furthermore, the positive coefficient aligns with the economic principle that when demand increases, price tends to increase. Since wages are a price of labor, they could rise as well in response to higher demand.

The input variable demonstrates a negative relationship with hourly wages. This could suggest that as industries sell more input to other industries, the hourly wages decrease. One explanation could be that firms, to maintain profit margins, might reduce their expenses, including wages, when they are selling more inputs (potentially at lower prices) to other industries. The negative coefficient suggests that an increase in worked hours is associated with a decrease in hourly wages. This could be due to various factors like overtime laws, which may not compensate for additional hours on a one-to-one basis, or the law of diminishing returns where additional hours of work do not necessarily result in proportional output. This aligns with the assumption that worked hours are determined by the demand for labor in the industry and can influence wages.

Table 5.4 presents the estimates for the Pooled Ordinary Least Square (POLS) regression model, with the ’Log of Hourly Wage’ as its dependent variable. There’s different potential issues that occurs for the given model, and we will therefore interpret those issues. We can firstly see, that the model appears to have a very high R-squared value at 0.9786, which indicates that 97.86% of the variance in the dependent variable ’Log of Hourly Wage’ can be explained with the used independent variable. While a great R-squared value is an indication that it explains the variation of the independent variables very well, but it also can also suggest, that there’s high correlation between the independent variables.

Hereunder we have, that we have included ’Log of Worked hours’ as an independent variable. Considering, that ’Log of Hourly wage’ are likely derived from ’Log of Total wages’ divided by ’Log of Worked hours’, this could cause a problem with endogeneity. The estimates for the coefficient could therefore be seen as biased and inconsistent, which will give a wrong picture, of how it really is.

Lastly we could have a problem with selecting the correct variables for our dependent variable. When selecting which variables should be estimated with regards to our dependent variable, it’s important to chose the right ones. We can’t just pick a lot of variables, and hope for the best. The reason for this is, that the model therefore will be ’overfitted’, and therefore it will estimate incorrect coefficient estimates and inflated standard errors. It can therefore also make the model more complex, and also harder to interpret. This can lead to incorrect conclusion between the dependent and independent variables.

6 Emperical Analysis

Model selection

In the following section, our primary objective is to identify the best-fitting econometric model for our data analysis, specifically comparing the POLS model and the FE model, and to determine whether the model are giving causal estimates through unbiasedness and consistent estimates. To achieve this, we will thoroughly examine each model's underlying assumptions, assess their performance on key indicators, and evaluate the extent to which they address potential econometric issues such as bias and consistency. POLS and FE models offer different approaches to panel data analysis, and selecting the most appropriate model for our study is crucial for drawing accurate and meaningful conclusions.

POLS

When working with Pooled OLS regressions certain assumptions needs to hold. We previously stated these assumptions and the reason we include this, is to get an unbiased and consistent estimator. For the Pooled OLS estimator to be unbiased and consistent, we need to hold assumptions MLR.1 through MLR.4. While the rest assumptions are still important for efficiency, they do not necessary lead to bias when violated.

MLR.1 holds as the model is linear in parameters. The model is flexible and allow for transformations within the parameters, but the relationship between the variables of interest is linear. When modelling for hourly_wage using Pooled OLS we get the specific model written as:

$$\begin{aligned} Hourly\_wage\_LN = {} & \beta_0 production\_LN + \beta_1 production\_cost\_LN + \beta_2 Germany\_imports\_LN \\ & + \beta_3 input\_LN + \beta_4 worked\_hours\_LN + u \end{aligned} \tag{6.1}$$

And when modelling for Employed_LN using pooled OLS we get the specific model:

$$\begin{aligned} Employed\_LN = \beta_0 production\_LN + \beta_1 production\_cost\_LN + \beta_2 Germany\_imports\_LN \\ + \beta_3 input\_LN + \beta_4 total\_wages\_LN + u \end{aligned} \tag{6.2}$$

We assume that MLR.2 holds as we collect data from 'Danmarks statistik' and 'stats.OECD', which provide insight to the economics of each industry. We collect impartial economic data for each industry, such as total wages, total worked hours by wage earners, total wage earners, production, production cost and imports to Germany. We assume that these sites are collecting their data through reliable sample of the population, such that we use data that represent the real world. If MLR.2 were violated, our model would be biased as we wouldn't use data that represent the real world. If the dataset is biased, the estimates could be biased.

MLR.3 holds as our variables of interest has no exact linear relationships. We allow for collinearity to some degree, as multiple regression analysis would be very limited if we assume no correlation between independent variables. We have not included any constant variable to our model, nor have we included any perfect linear transformations. If MLR.3 is violated we would see this when interpreting the estimation results. When including perfect linear transformations, the variables in question would not be able to perform as needed in the model, as we changes one and hold the other fixed. The perfect linear transformation would not be fixed, and the estimates would be biased.

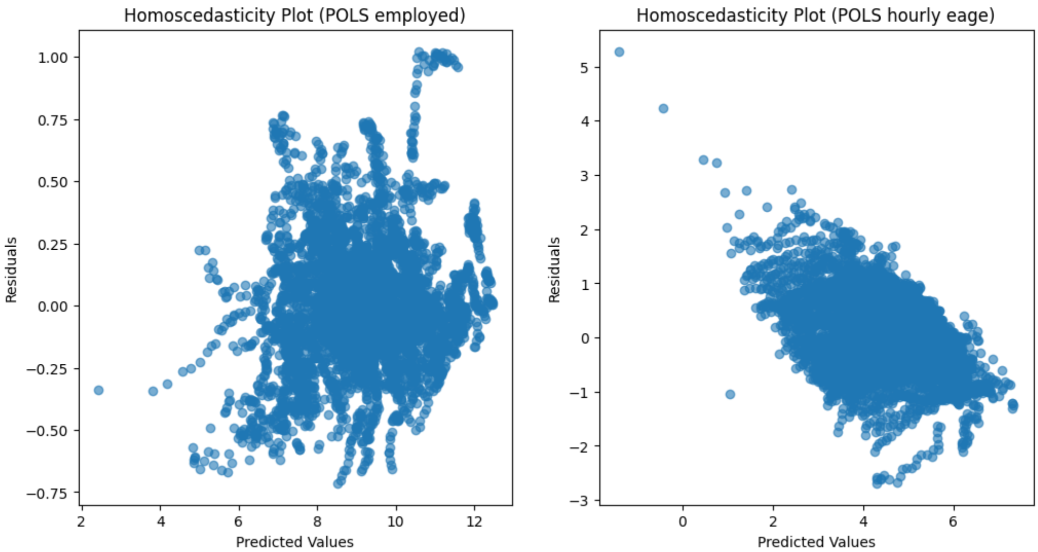
MLR.4 is the most important assumption to ensure unbiasedness, and may cause us problems due to our dataset. We are working with panel data, where the main idea is to highlight the unobserved effects. The assumption dictates that the error term has an expected value of zero, given any values of the independent variables. The error term (u_{it}) represents the unobserved factors or disturbances that affect the dependent variable (y_{it}) but are not captured by the independent variables ($X_{it1}, X_{it2}, \cdots, X_{itk}$). These factors can include omitted variables, measurement errors, or other unobserved determinants of the outcome.

$$E(u_{it}|X_{it1}, X_{it2}, \cdots, X_{itk}) = 0$$

Assuming MLR.4 allows us to use ordinary least squares (OLS) estimation, which provides unbiased estimates of the coefficients and enables valid statistical inference, such as hypothesis testing and confidence intervals

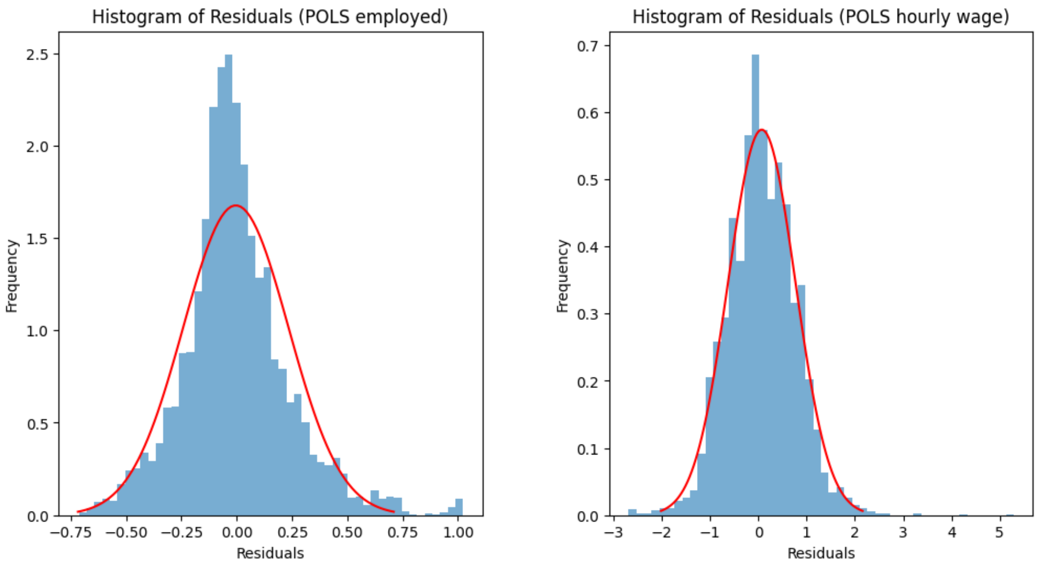
MLR.5 holds as the data indicates constants variance for residuals in each industry and time period, leading to homoskedasticity. The variance of each of the error terms is constant, meaning the variance over time is constant. When plotting the residuals against the predicted values, we see that the scatter plot appears not to be spreading out but more dense with constant variance. We do experience some outliers, but the overall look is looking somewhat homoskedastic.

Figure 6.1: predicted values vs residuals, scatterplot for employed and hourly wage



MLR.6 seems to hold as the residuals are somewhat normal distributed, $N(0, \sigma)$. There is some skewness when examining the model for employed as explained variable, but when looking at both distributions they seems likely to be normal distributed.

Figure 6.2: distribution of residuals compared to normal(0, σ^2), histogram for employed and hourly wage



Fixed Effects

When working with the Fixed Effects estimator we need the data to behave in a certain way, just like with the Pooled OLS. the main difference between the POLS and the FE estimator, is that the FE estimator differences out the Fixed Effects (time-invariant) before using the OLS. For the Fixed Effects estimator to be unbiased and consistent, we need to hold assumptions FE.1 though FE.4. While the rest assumptions are still important, they do not necessary lead to bias when violated.

FE.1 holds as the model is linear in parameters. Like with the pooled OLS the model is flexible and the underlying variables can be transformations like natural logarithms, squared etc. When modelling for hourly_wage_LN using Fixed Effects we get the specific model written as:

$$\begin{aligned} Hourly\_wage\_LN_{it} = \beta_0 production\_LN + \beta_1 wage\_earners\_LN_{it}\beta_2 production\_cost\_LN_{it} \\ + \beta_3 total\_wages\_LN_{it} + a_i + u_{it} \end{aligned} \tag{6.3}$$

And when modelling for Employed_LN using Fixed Effects we get the specific model:

$$\begin{aligned} Employed\_LN_{it} = \beta_0 production\_LN + \beta_1 Germany\_imports\_LN_{it} + \beta_2 production\_cost\_LN_{it} \\ + \beta_3 input\_LN_{it} + \beta 4 worked\_hours\_LN + a_i + u_{it} \end{aligned} \tag{6.4}$$

Using the Fixed Effects estimator we include the unobserved effect in the model, before the model differences out the time-invariant effects, by subtracting the average of the equation from each year.

We assume that FE.2 holds as we collect data from 'Danmarks statistik' and 'stats.OECD', which provide insight to the economics of each industry. We collect impartial economic data for each industry, such as total wages, total worked hours by wage earners, total wage earners, production, production cost and imports to Germany. We assume that these sites are collecting their data through reliable sample of the population, such that we use data that represent the real world. If FE.2 is violated, our model would be biased as we wouldn't use data that represent the real world. If the dataset is biased, the estimates are biased.

FE.3 holds as all variables changes for some i. Some of the variable could not be filled with data, but we have not included constant variables. We do not have perfect linear relationships between the variables of interest, but again we have some correlation. if FE.3 is violated the estimates may be biased, we therefore choose our variables wisely.

FE.4 holds and the expected value of the residuals given any explanatory variable, X_i, and the unobserved effect, a_i, is zero. In a Fixed Effects model, the unobserved individual-specific effect (a_i) is introduced to capture time-invariant heterogeneity among individuals. By including individual fixed effects, we are essentially differencing out all time-invariant factors, such as individual-specific preferences, or characteristics, which are constant over time.

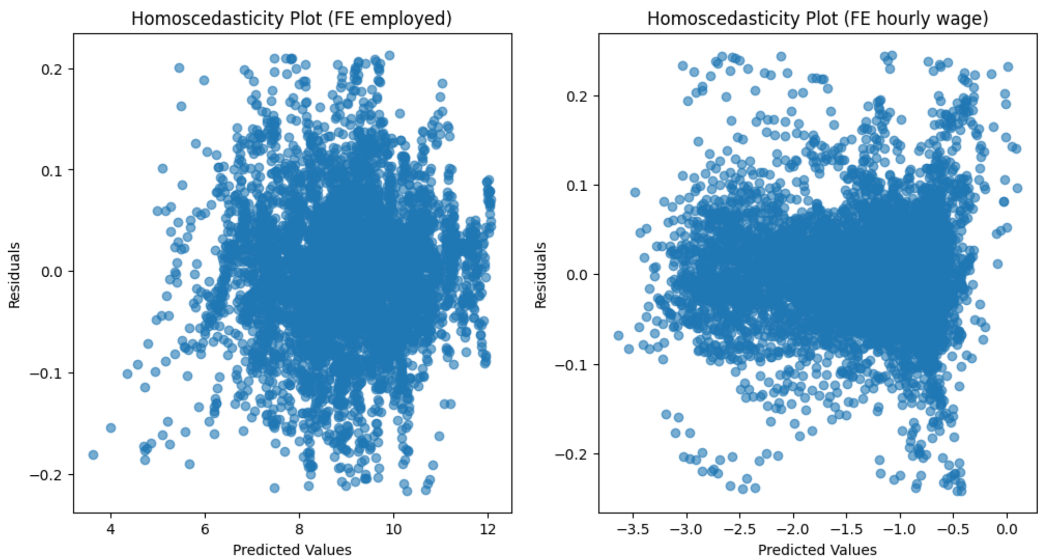
$$E(u_{it}|X_{it}, a_i) = 0$$

The exogeneity assumption states that the error term (u_{it}) is uncorrelated with the explanatory variables (X_{it}). In the fixed effects model, we extend this assumption to include the unobserved effect (a_i) as well. This means that the error term is uncorrelated with both observed and unobserved factors. By assuming strict exogeneity, we can consistently estimate the fixed effects model and obtain valid statistical inference. This assumption enables us to differentiate the effects of observed variables from the effects of unobserved individual-specific factors on the outcome variable.

FE.5 holds as the industry and time period individual idiosyncratic error term has a constant variance of zero, which leads to homoskedasticity. Homoskedasticity leads to residuals not spreading when plotted against the predicted values. We filtered out the large outlier and the plot seems homoskedastic.

$$Var(u_{it}|X_i, a_i) = Var(u_{it}) = \sigma^2$$

Figure 6.3: predicted values vs residuals, scatterplot for employed and hourly wage



FE.6 holds as all idiosyncratic errors, $t \neq s$, are uncorrelated, given all explanatory variables and unobserved effect. We found that the covariance between u_{it} and u_{is} across all industries and time periods is zero:

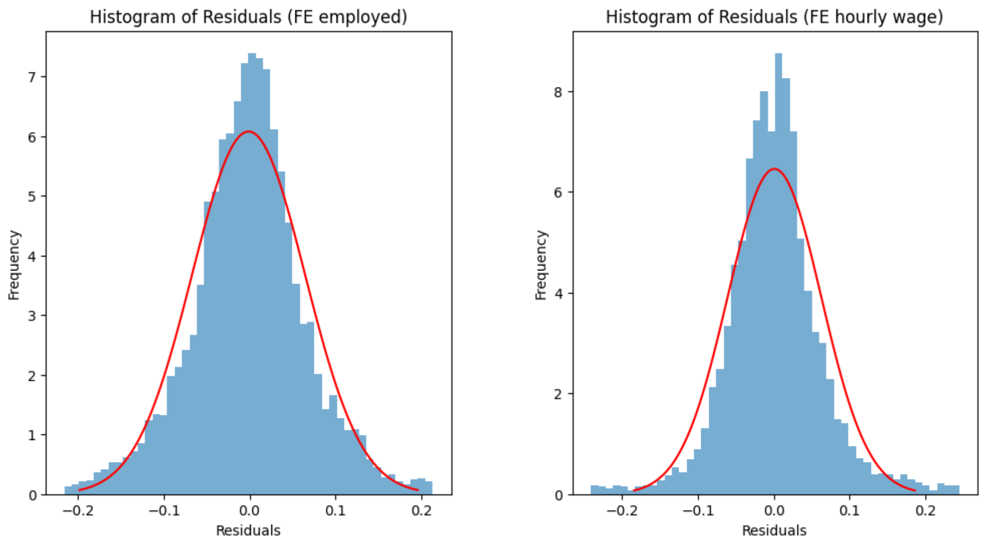
$$Cov(u_{it}, u_{is}|X_i, a_i) = 0$$

Under FE.1 through FE.6 we have the best linear unbiased estimator (BLUE). The Best Linear Unbiased Estimator is a concept in statistics that refers to an estimator that satisfies two important properties: linearity and unbiasedness.

–In linear regression, the goal is to estimate the relationship between a dependent variable and one or more independent variables–. The BLUE is the estimator that achieves the minimum variance among all linear unbiased estimators. The BLUE is optimal in the sense that it has the smallest variance among all linear unbiased estimators. In other words, it provides the most precise and efficient estimate of the unknown parameters. The estimator is linear because it is a linear combination of the observed data. It can be expressed as a weighted sum of the explanatory variables, where the weights are chosen to minimize the variance of the estimator. The estimator is unbiased when its expected value equals the true value of the parameter being estimated. In other words, on average, the estimator does not systematically overestimate or underestimate the parameter.

FE.7 holds as all the idiosyncratic errors, conditional on the explanatory variables, roughly follows the normal distribution. We do experience some large outliers, but the distribution becomes clear when filtered out. This implies FE.4 through FE.6 but is stronger, as it assumes normal distribution for the idiosyncratic errors.

Figure 6.4: distribution of residuals compared to normal(0, σ^2), histogram for employed and hourly wage



Model specification

Selecting the appropriate variables for our model is crucial in order to obtain the most accurate and meaningful results. It is essential to consider the relative importance of each variable, particularly in relation to our main research question.

The process of variable selection requires careful consideration and analysis. By including variables that carry significant information related to our main question, we can enhance the predictive power and explanatory capacity of the model. These variables should provide relevant insights and contribute to a comprehensive understanding of number of employed and the hourly wages in each industry. It is important to acknowledge that not all variables are equally informative or influential in our analysis. Some variables may carry more weight and contribute more significantly to the outcomes we are interested in. Identifying and prioritizing these influential variables is crucial to ensure the model's effectiveness.

In addition to considering the significance of each variable, we should also evaluate potential interactions or relationships between variables. Certain variables may interact with one another, amplifying or diminishing their individual effects. By including such interaction terms, we can capture more nuanced relationships and improve the model's accuracy. Furthermore, the selection process should involve a rigorous evaluation of the variables' reliability and relevance. It is important to exclude any variables that are redundant, have limited variability, or lack a strong theoretical or empirical basis. These variables may not contribute substantially to the model's performance and may even introduce noise or bias. During the model selection process, we take into account multiple factors to ensure we choose the best possible model. These factors include the significance of coefficients, goodness of fit measures such as R-squared and the F-statistic, as well as information criteria like AIC and BIC. To begin the model selection process, we typically start with including one independent variable and examine the corresponding p-values. By assessing the significance of the coefficient associated with that variable, we can determine its individual impact on the dependent variable. This step allows us to build a foundation for our model and understand the initial relationships between variables.

As we progress, we gradually introduce additional independent variables, carefully considering their p-values and their significance in relation to the research question at hand. Including variables one by one helps us assess their individual contributions to the model and identify the most meaningful predictors. After including each variable, we closely evaluate the goodness of fit measures such as R-squared and the F-statistic. These measures provide valuable insights into how well our model fits the data and captures the variance in the dependent variable. Higher values of R-squared and significant F-statistic indicate a better fit and stronger explanatory power. Additionally, we consider information criteria such as AIC and BIC. These criteria account for both the goodness of fit and the complexity of the model, preventing overfitting by penalizing the inclusion of unnecessary variables. Lower values of AIC and BIC indicate a more parsimonious model that still captures the essential information in the data.

Throughout the model selection process, we compare different models based on these factors, taking into account the significance of coefficients, goodness of fit measures, and information criteria. This allows us to make informed decisions about which variables to include and which model to choose as the best representation of the data and the most suitable for answering the research question. By combining rigorous statistical analysis with a thoughtful consideration of these factors, we can build robust models that provide accurate and meaningful insights, ensuring that our research findings are reliable and valid.

When selecting explanatory variables for the employed, it is important to carefully consider the inclusion of wage earners. Since wage earners represent a similar concept, it may not be ideal to include as an explanatory variable in the model. Regarding models for employed, Including both total wages and worked hours alone in the models results in a high R-squared values suggesting that the other variables may lose their significance or explanatory power. We may be restricted to only include one or even not include either of total wages and worked hours. In models for hourly wages, Since the hourly wage is directly derived from $\frac{total\ wages}{worked\ hours}$ it may not be ideal to include both as explanatory variables in the model. Since wage earners and employed represent a similar concept, carefully consideration of which to include. Production is a crucial factor as it reflects economic activity and can be influenced by both external shocks and labor market conditions. Germany imports can provide insights into the international trade dynamics and its potential impact on the domestic labor market. Production costs and inputs, on the other hand, can capture the cost pressures faced by firms and shed light on the relationship between labor market conditions.

Fixed effect - Employed

Table 6.1

Dependent variable: Log of Employed						
Model	(1) FE	(2) FE	(3) FE	(4) FE	(5) FE	(6) FE
	Coefficient (p-value) [std.err.]	Coefficient (p-value) [std.err.]	Coefficient (p-value) [std.err.]	Coefficient (p-value) [std.err.]	Coefficient (p-value) [std.err.]	Coefficient (p-value) [std.err.]
Log of production	0.8679 (0.0000) [0.0075]	0.8269 (0.0000) [0.0073]	0.9196 (0.0000) [0.0175]	0.9254 (0.0000) [0.0176]	0.0989 (0.0000) [0.0097]	0.0404 (0.0000) [0.0061]
Log of Germany import	-	-0.0267 (0.0000) [0.0011]	-0.0273 (0.0000) [0.0011]	-0.0272 (0.0000) [0.0011]	-0.0110 (0.0000) [0.0005]	-0.0040 (0.0000) [0.0003]
Log of Production cost	-	-	-0.1002 (0.0000) 0.0172	-0.0947 (0.0000) [0.0173]	0.0168 (0.0328) [0.0079]	0.0400 (0.0000) [0.0050]
Log of Input	-	-	-	-0.0106 (0.0004) [0.0030]	-0.0192 (0.0000) [0.0014]	-0.0079 (0.0000) [0.0009]
Log of Total wages	-	-	-	-	0.8624 (0.0000) [0.0057]	-
Log of Worked hours	-	-	-	-	-	0.8805 (0.0000) [0.0034]
R^2	0.6895	0.7184	0.7200	0.7205	0.9423	0.9770
F-statistics	1.341e+04	7704.1	5175.4	3892.1	1.971e+04	5.138e+04
F-test		1.11e-16	3.74e-09	0.00032	1.11e-16	1.11e-16
Log-likelihood	-568.33	-264.84	-247.45	-240.97	4655.7	7518.5
AIC	1138.66	533.68	500.90	489.94	-9301.30	-15026.99
BIC	1145.40	547.15	521.106	516.88	-9267.63	-14993.32

Comparing the coefficient significance, we observe that in model 1 through model 6, the coefficient for the independent variables are highly significant with p-values bellow our significance level of 0.05(5%). We cannot exclude any variables, at any step from the model.

When including $production\_LN$ as the only variable, we find that the model yields quite good results. When including more independent variables, we increase the model fit end gain a better model. When we include $total\_wages\_LN$ and $worked\_hours\_LN$, the model fit is almost perfect.

When examining the goodness of fit, the R^2 values indicate the proportion of variance explained by each model. We clearly see a change when individually including $total\_wages\_LN$ and $worked\_hours\_LN$, $R^2 = 0.9423(94, 23\%)$ and $R^2 = 0.9770(97, 70\%)$ respectively. This could be leading to shift a in the explanatory power of all other variables, which we really want to estimate.

The F-statistics assess the overall significance of the models. Model 1 through model 6 has high f-statistics, meaning high significance. The f-test indicates a better model then the one to the left, as they are very close to asymptotic zero(for model 5 and model 6, we compare to model 4).

Considering the information criteria, lower values of AIC and BIC indicate better model fit. Model 6 has the lowest AIC and BIC values of approximately -15026 and -14993, respectively. We see that, as we include more dependent variables the AIC and BIC grow smaller and smaller. We see quite large value changes from model 4 to model 5 and again to model 6, leading us to believe that the model becomes better.

Based on our model selection analysis, model 6 emerges as the best choice. It exhibits the highest R-squared value, indicating a better fit to the data, and also has a highly significant coefficient for the independent variable. Additionally, Model 6 has the highest log-likelihood and the lowest AIC and BIC values, further supporting its selection as the superior model.

Fixed effect - Hourly wage

Table 6.2

Dependent variable: Log of Hourly Wage						
Model	(1) FE	(2) FE	(3) FE	(4) FE	(5) FE	(6) FE
	Coefficient (p-value) [std. err.]	Coefficient (p-value) [std. err.]	Coefficient (p-value) [std. err.]	Coefficient (p-value) [std. err.]	Coefficient (p-value) [std. err.]	Coefficient (p-value) [std. err.]
Log of Production	-0.0152 (0.0000) [0.0033]	0.0746 (0.0000) [0.0057]	0.0826 (0.0000) [0.0055]	0.0731 (0.0000) [0.0054]	-0.0169 (0.0137) [0.0069]	0.0769 (0.0000) [0.0090]
Log of Wage earners	-	-0.1015 (0.0000) [0.0053]	0.3785 (0.0000) [0.0214]	0.4383 (0.0000) [0.0214]	-0.7431 (0.0000) [0.0093]	0.4739 (0.0000) [0.0213]
Log of Worked hours	-	-	-0.4811 (0.0000) [0.0208]	-0.5173 (0.0000) [0.0206]	-	-0.5509 (0.0000) [0.0205]
Log of Germany import	-	-	-	0.0071 (0.0000) [0.0005]	-	0.0070 (0.0000) [0.0005]
Log of Production cost	-	-	-	-	0.0168 (0.0027) [0.0056]	-0.0248 (0.0007) [0.0073]
Log of Total wages	-	-	-	-	0.7344 (0.0000) [0.0097]	-
Log of Input	-	-	-	-	-	0.0176 (0.0000) [0.0013]
R^2	0.0034	0.0607	0.1368	0.1667	0.5196	0.1931
F-statistics	20.832	195.20	319.08	302.06	1632.6	240.76
F-test	-	1.11e-16	1.11e-16	1.11e-16	1.11e-16	1
Log-likelihood	4466.4	4650.2	4912.6	5022.1	6732.0	5122.0
AIC	-8930.86	-9296.43	-9819.17	-10036.24	-13456.03	-10231.93
BIC	-8924.13	-9282.96	-9798.97	-10009.31	-13429.10	-10191.52

Comparing the coefficient significance, we observe that in model 1 through model 6, the coefficient for the independent variables are highly significant with most of the p-values=0. We have largest p-value=0.0137 which is well below our chosen significance level of 0.05 (5%).

Estimating with only $production\_LN$ yields very poor results, but as we add more explanatory variables the model yield better results. The explanatory variables does not yields very good results, but we get two estimations the may be of use. One model include all variables but

$total\_wages\_LN$, leading us to have coefficients which we can use in our main research question. The other include $total\_wages\_LN$ and yields a better fit, but fewer estimated coefficients.

When examining the goodness of fit, model 5 has the highest R^2 value of 0.5196(51.95%). Model 6 show bad R^2 value of 0.1931(19.31%), leading us to believe model 5 to be the best.

The F-statistics assess the overall significance of the models. Model 1 through model 6 has high f-statistics, meaning high significance. The f-test indicates a better model then the one to the left, as they are very close to asymptotic zero.

Considering the information criteria, model 6 has the lowest AIC and BIC values of approximately -13456 and -13429, respectively. This is a significant change when compared to the next best.

Model 5 emerges as the best choice, mainly due to best values. It exhibits the highest R-squared value, indicating a better fit to the data. While it is not a great fit, it is far better than the other, leading ud to believe they med suffer from omitted vaiables. Model 5 also has a highly significant coefficient for the independent variables. Additionally, model 5 has the highest log-likelihood and the lowest AIC and BIC values, further supporting its selection as the superior model.

POLS - Employed

Table 6.3

Dependent variable: Log of Employed						
Model	(1) POLS	(2) POLS	(3) POLS	(4) POLS	(5) POLS	(6) POLS
	Coefficient (p-value) [std.err.]	Coefficient (p-value) [std.err.]	Coefficient (p-value) [std.err.]	Coefficient (p-value) [std.err.]	Coefficient (p-value) [std.err.]	Coefficient (p-value) [std.err.]
Log of production	1.0531 (0.0000) [0.0018]	-	-	1.4850 (0.0000) [0.0168]	1.0296 (0.0000) [0.0265]	0.0456 (0.0000) [0.0055]
Log of Production cost	-	1.1448 (0.0000) [0.0029]	-	-0.4105 (0.0000) [0.0193]	-0.3253 (0.0000) [0.0190]	0.0099 (0.0240) [0.0044]
Log of Germany import	-	-	-0.1718 (0.0000) [0.0143]	-0.0742 (0.0000) [0.030]	-0.0661 (0.0000) [0.0029]	-0.0049 (0.0000) [0.0007]
Log of Input	-	-	1.13661 (0.0000) [0.0137]	-0.0595 (0.0000) [0.0052]	-0.0579 (0.0000) [0.0050]	-0.0223 (0.0000) [0.0011]
Log of Total wages	-	-	-	-	0.4399 (0.0000) [0.0203]	-
Log of Worked hours	-	-	-	-	-	0.9314 (0.0000) [0.0026]
R^2	0.9813	0.9628	0.6878	0.9862	0.9993	0.9872
F-statistics	3.26e+05	1.606e+05	6839.1	1.112e+05	1.901e+06	9.578e+04
F-test			1.0	1.11e-16	1.11e-16	1.11e-16
Log-likelihood	-1.032e+04	-1.246e+04	-1.906e+04	-9367.2	100.00	-9139.7
AIC	20635.95	24913.86	38110.61	76.84	-190.00	18289.39
BIC	20642.68	24920.59	38130.81	97.04	-156.33	18323.05

Comparing the coefficient significance, we observe that in model 1 through model 6, the coefficient for the independent variables are highly significant with most p-values of zero, and just a few well below our chosen significance level of 0.05 (5%).

When including $production\_LN$ and $production\_cost\_LN$ as the only variable, we find that both the model yield very good fit. When including $Germany\_imports\_LN$ and $input\_LN$ in one model, we get a fine model but not as good fit. when Estimating model 4, we include all four variables and we reach the best model yet. Model 5 and model 6 include $total\_wages\_LN$ and $worked\_hours\_LN$ respectively, and yields slightly higher fit compared to model 4.

When examining the goodness of fit, the R^2 values indicate the proportion of variance explained by each model. we clearly see a good fit on all models, and that model 5 has the highest R^2 value of 0.9993 (99.93%). The R^2 support the selection of model 5.

The F-statistics assess the overall significance of the models. Model 1 through model 5 has high f-statistics, meaning high significance. The f-test indicates that model 4 is better than model 1 and model 2 as it is close to asymptotic zero. Model 5 and model 6 are better model than model 4, but they include $total\_wages\_LN$ and $worked\_hours\_LN$ which could lead to the other variables may lose their significance or explanatory power.

Considering the information criteria, lower values of AIC and BIC indicate better model fit. Model 5 has the lowest AIC and BIC values of approximately -190 and -156, respectively. This is a significant change when compared to the next best.

Based on the model selection analysis, Model 5 emerges as the best choice. It exhibits the highest R-squared value, indicating a better fit to the data, and also has a highly significant coefficient for the independent variable. Additionally, Model 5 has the highest log-likelihood and the lowest AIC and BIC values, further supporting its selection as the superior model.

POLS - Hourly wage

Table 6.4

Dependent variable: Log of Hourly wages						
Model	(1) POLS	(2) POLS	(3) POLS	(4) POLS	(5) POLS	(6) POLS
	Coefficient (p-value) [std.err.]	Coefficient (p-value) [std.err.]	Coefficient (p-value) [std.err.]	Coefficient (p-value) [std.err.]	Coefficient (p-value) [std.err.]	Coefficient (p-value) [std.err.]
Log of production	0.5385 (0.0000) [0.0011]	0.5818 (0.0000) [0.0118]	0.6253 (0.0000) [0.00113]	0.6169 (0.0000) [0.0115]	0.6667 (0.0000) [0.0197]	0.9635 (0.0000) [0.0161]
Log of Production cost	-	-0.0478 (0.0002) [0.0130]	-0.1138 (0.0000) [0.0125]	-0.0956 (0.0000) [0.0132]	-0.0908 (0.0000) [0.0142]	-0.1968 (0.0240) [0.0129]
Log of Germany import	-	-	0.0503 (0.0000) [0.0019]	0.0538 (0.0000) [0.0021]	-	0.0371 (0.0000) [0.0020]
Log of Input	-	-	-	-0.0156 (0.0000) [0.0035]	0.0191 (0.0000) [0.0034]	-0.0245 (0.0000) [0.0033]
Log of Total wages	-	-	-	-	-0.0673 (0.0000) [0.0159]	-
Log of Worked hours	-	-	-	-	-	-0.2243 (0.0000) [0.0078]
R^2	0.9728	0.9729	0.9756	0.9757	0.9731	0.9786
F-statistics	2.221e+05	1.113e+05	8.278e+04	6.227e+04	5.613e+04	5.667e+04
F-test		0.00023	1.11e-16	1.09e-05	1.0	1.11e-16
Log-likelihood	-7343.5	-7336.7	-7004.8	-6995.1	-7309.3	-6603.2
AIC	14688.92	14677.36	13729.23	13998.17	14626.58	13216.47
BIC	14695.65	14690.83	13749.43	14025.11	14653.52	13250.13

Comparing the coefficient significance, we observe that in model 1 through model 5, the coefficient for the independent variables are highly significant with most of the p-values=0. We have one p-value=0.0240 which well below our chosen significance level of 0.05(5%).

When including $production\_LN$ as the only variable, we find that the model yields very good results. When including more independent variables, we increase the model fit end gain a slightly better model. When we initially include $total\_wages\_LN$ it is insignificant until we remove $Germany\_imports\_LN$.

When examining the goodness of fit, the R^2 values indicate the proportion of variance explained by each model. We clearly see a high fit on all of the models, and that model 6 has the highest R^2 value of 0.9786(97.86%). The R^2 is very close among all models, meaning no significant changes.

The F-statistics assess the overall significance of the models. Model 1 through model 6 has high f-statistics, meaning high significance. The f-test indicates that model and model 6 is the better model, as they are asymptotic very close to zero when compared with the model the the left. model 4 and model 5 are both compared to model 3, and though model 4 has a higher R^2 value, the overall model is worse with f-test value of 1.0.

Considering the information criteria, lower values of AIC and BIC indicate better model fit. Model 6 has the lowest AIC and BIC values of approximately 13216 and 13250, respectively. We see that, as we include more dependent variables the AIC and BIC converges and the changes grows smaller and smaller. The changes in AIC and BIC in each step is not as great as we have seen earlier, but still significant.

Based on our model selection analysis, model 6 emerges as the best choice. It exhibits the highest R-squared value, indicating a better fit to the data, and also has a highly significant coefficient for the independent variable. Additionally, Model 6 has the highest log-likelihood and the lowest AIC and BIC values, further supporting its selection as the superior model.

Omitted variable bias

Omitted variable bias occurs when a relevant variable is not included in a statistical model, which can lead to biased and inconsistent estimates. In our fixed effect models, we have controlled industry and time effects and therefore addressed potential bias. When we control for industry fixed effects, you control for all time-invariant industry-specific factors this means we control for rules and regulations that govern an industry can be time-invariant industry-specific factors. For example, some industries may be subject to more stringent environmental regulations, which can affect their production costs and hence their wages or other factors such as capital intensity, skill requirements, market structure and technological advancements. By including time effects, we control for any global shocks or trends that might affect all industries equally.

However, we must keep in mind even though we control for a lot to avoid bias there still could be unobserved factors that vary both over time and across industries. For instance, there may be a unique event or policy change that affects one industry in a particular year. This would not be captured by industry or time fixed effects. Essentially, it's like adding a dummy variable for each time period.

Let's start by taking a look of table 5.1, we can see total wages and worked hours were excluded because they are functions of each other. If these variables are perfectly (or near perfectly) correlated, including them could lead to multicollinearity, which can cause problems in estimating wrong coefficients and inflating standard errors for the model.

Referring to Table 5.2, we've decided to exclude 'Germany imports' from our models because of its insignificance. Similarly, 'worked hours' has also been omitted due to its lack of significance. Given that we no longer include 'worked hours' and that 'total wages' is function on 'worked hours', we are now in a position to incorporate 'total wages' into our model. Furthermore, we've chosen to exclude 'input', which was also found to be insignificant.

For Table 5.3, we have the same scenario as Table 5.2. 'Worked hours' has been excluded, as 'worked hours' and 'total wages' can't both be included as explanatory variables, as they are a derived function of each other. We have therefore excluded 'worked hours', but included 'total wages' as it made better sense.

Drawing from Table 5.4 we exclude worked hours has been excluded due to its marginal impact. As 'worked hours' is no longer included and considering 'total wages' is contingent on 'worked hours', we are able to include 'total wage' in our model. Additionally, 'input' has reached a statistical significance that warrants its inclusion.

Partial conclusion

Panel data analysis with industry and time included often exhibits correlation within each industry. In such cases, the fixed effects (FE) method is more suitable, especially when observing multiple industries over time. FE models are particularly useful when the primary interest lies in understanding the variation within industries over time rather than comparing across industries.

The main benefit of using the FE method is its ability to account for time-invariant unobserved factors that are constant within industries. By including industry-specific fixed effects, these unobserved factors can be effectively eliminated. Moreover, FE models tend to be more efficient than the pooled ordinary least squares (POLS) models since they leverage the within variation of each individual industries.

While POLS regression is consistent and unbiased under assumption MLR.1 through MLR.4, it can be biased if unobserved effects exist, resulting in consistently over- or underestimated estimates. POLS models are appropriate for cross-sectional data, where different industries are observed at a single point in time. They assume that the coefficients are constant across industries, assuming no systematic industry -specific heterogeneity. However, as the correlation between industries and time becomes evident in the data, the assumptions of POLS may not hold.

Compared to POLS models, FE models are superior in our case and data for several reasons. Firstly, the FE estimator is best linear unbiased estimator (BLUE), providing more reliable estimates. Additionally, FE models account for the within-entity variation, making them more efficient. Hence, considering the systematically correlated nature of the data, the FE method is preferred in our analysis.

7

Discussion

Choosing the data

Since we have chosen panel data there are several advantages over both cross-sectional data and time series data, especially when examining the impacts of supply and demand shocks on labor market. Panel data allows for the modeling of effects that are both time-invariant and individual-specific. For instance, there might be industry-specific characteristics that don't change over time that affect both wages and employment. Cross-sectional data and time series data separately cannot account for these factors simultaneously. With panel data, we can control for variables that we cannot observe or measure. By looking at changes within an entity over time, we can control for time-invariant unobserved heterogeneity. This is not possible with pure cross-sectional or time series data. Panel data allows us to analyze dynamics of change, such as lags in behavior or adaptation to changes (such as shocks). For instance, it might take time for the effects of a demand shock to fully filter through to wages and employment, something you can model with panel data. Panel data account for interactions between variables over time. For instance, how changes in supply and demand shocks influence changes in wages and employment.Time-series data allows for the analysis of trends over time, but it doesn't let us compare across different industries. It would limit the analysis to how the variables change over time in the whole economy or in a single industry. Cross-sectional data, on the other hand, allows for comparisons across different industries, but without considering how these conditions change over time.

Our initiate approach to the research question was to choose the variables which seemed to inform us about the employment and hourly wages over time. We deemed the variables relevant and enough to perform our analysis. First, let's take a closer look at the 'wage earners' variable. This simply stands for the number of people in an economy who are earning wages. It's a big deal because it influences both hourly wages and the number of people employed. Our panel data model allows us to see how the number of wage earners changes over time, giving us valuable insight into whether the labor market is growing or shrinking. This also gives us a better understanding of how changes in supply and demand affect the labor market. Next up is 'total wages.' This is just the total amount of wages paid out in the economy. It's like a snapshot of the overall wage levels, and it's really useful. We use panel data to track changes in total wages over time and across different areas. Doing this helps us see how these changes align with shifts in employment levels and hourly wages. We also look at 'worked hours,' which is an important way to measure labor supply. Our panel data model allows us to see how changes in worked hours, which might suggest changes in labor market conditions, affect wages and employment over time. We then turn our attention to the 'production' variable. This refers to the amount of production in an economy or a specific industry, and it's a key sign of economic activity. With panel data, we can see how changes in production levels across different industries and over time result in adjustments in labor. These adjustments can then lead to changes in wages and employment rates. When we consider the 'input' variable, which represents the resources used in production. The 'input' variable in our study lets us look closely at how changes in the level or cost of inputs can affect production costs, wages, and job rates. Using panel data, we can track these impacts over time and across various sectors of the economy. We also keep an eye on 'production cost.' Being able to see how production costs can influence wages and job rates, especially when supply and demand change, is a big reason we chose to use panel data. Finally, we check out 'Germany imports.' The amount of imports can give us a good idea of how well an economy is doing, as it shows the demand for goods from other countries. With panel data, we can see how changes in the amount of imports can affect wages and job rates over time and in different industries.

We could have expanded our dataset such that it could provide additional insight. Here are a few potential variables that could be added. One example could be experience. Here the level of experience of the workforce in a given industry can significantly impact how the industry reacts to supply and demand shocks. For instance, more experienced workers might be more adaptable or productive, mitigating the impact of such shocks. Age can be another important factor. Older workers might have more experience and skills, but younger workers might be more adaptable to change. The age distribution of an industry's workforce can influence how it responds to shocks. Moreover, the educational qualifications of the workforce can play a role in determining the resilience of an industry to shocks. Higher levels of education can indicate a more skilled workforce, potentially influencing the response to changes in supply and demand. The location of the industry can impact the effects of supply and demand shocks. Some regions might have more robust supply chains, access to markets, or other factors that mitigate or amplify shocks. Also the level of skills in specific areas (such as technology skills, or specific vocational skills) can be crucial. Industries with a higher skill level may be better positioned to adapt to changes and mitigate the impacts of shocks.

We didn't expand upon our dataset, due to the data not being available. But a few more beneficial reasons could be, when adding more variables it will increase the risk of multicollinearity, where two or more variables are highly correlated. This can make it difficult to determine the individual effects of each variable on the dependent variable, leading to unstable estimates in our model. Moreover, the higher the risk of overfitting. Lastly, each additional variable adds a new dimension to the analysis, making the results more complicated to interpret. It can be challenging to effectively communicate the findings of a model with a large number of variables. This doesn't mean that additional variables are never useful, but their inclusion should be carefully considered in light of these trade-offs.

Strengths and limitations

Pooled Ordinary Least Squares (POLS) and Fixed Effects (FE) models represent two viable options, but their suitability depends on the assumptions we make about the data and their individual strengths and limitations. The Pooled-OLS model treats the dataset as a simple cross-section, pooling data across all industries and years, and neglects the panel structure of the data. This model operates on the assumption that there aren't any specific unobserved effects associated with different industries that are invariant over time. If these unobserved effects exist and are correlated with the independent variables, the Pooled-OLS model is likely to exhibit omitted variable bias. Despite its simplicity and capability to estimate effects of time-invariant variables, Pooled-OLS has major drawbacks. If we are dealing with panel data that features unobservable effects specific to different industries or time periods, Pooled-OLS could lead to inefficient and biased estimates. Moreover, the Pooled-OLS model assumes that error terms are uncorrelated across time and industries. Violations of these assumptions could result in incorrect standard errors, potentially leading to misleading inferences.

On the other hand, the Fixed Effects model, which is tailored specifically for panel data, offers a more nuanced analysis by controlling for entity-specific effects. For instance, in this study, the FE model would account for any unobservable, industry-specific characteristics that are constant over time but might influence wages and employment. By doing so, the FE model mitigates the risk of omitted variable bias, a major strength over the Pooled-OLS model. Another notable advantage of the FE model is its utilization of within-entity variation. It explores relationships between predictor and outcome variables within an entity (an industry, in this context), removing the effect of time-invariant characteristics. This facilitates an assessment of the net effect of predictors (e.g., supply and demand shocks) on the outcome variables (hourly wage and employment). However, the Fixed Effects model is not without its own limitations. It cannot estimate the effects of time-invariant predictors. Since the FE model is built upon changes within entities, any variable that doesn't change over time is omitted. Furthermore, the FE model assumes that all time-varying variables are exogenous, meaning they are not affected by the unobserved factors captured by the fixed effects. If there are endogenous variables or feedback effects between employment, wages, and the shocks the FE model may not provide unbiased estimates. Given the panel data structure in this study, incorporating data from multiple industries observed over several years, and potential existence of industry-specific characteristics that could influence the relationship between supply and demand shocks and the outcome variables, the Fixed Effects model is likely to be more suitable. This model allows us to control for these unobserved characteristics and provides more accurate and reliable estimates than the Pooled-OLS model.

Other approaches

The Random Effects would be a valuable regression model due to its particular handling of individual-specific effects. In our case, these are the industry-specific effects which are not directly observed but potentially play a crucial role in determining both wages and employment levels. The RE model assumes that these unobserved, industry-specific effects are random and uncorrelated with the regressors (independent variables), such as supply and demand shocks. By treating these effects as random components of the error term, the RE model allows for the estimation of time-invariant variables, which the Fixed Effects (FE) model cannot do. For instance, if there were industry characteristics that do not change over time, the RE model would be capable of capturing the impact of these characteristics on wages and employment. Furthermore, the RE model utilizes both within-entity (industry) and between-entity variations, making it more efficient than the FE model under the assumption that unobserved, industry-specific effects are not correlated with the independent variables. This means, if the assumptions hold, the RE model could provide more accurate estimates with less variance compared to the FE model. However, it's crucial to remember that these benefits hinge on the model's assumptions being valid. If the unobserved effects are not random but correlated with the regressors, using the RE model can lead to biased and inconsistent estimates. In such scenarios, the FE model would be more suitable. In the context of this study - investigating the impacts of supply and demand shocks on hourly wages and employment - it is quite plausible that unobserved industry-specific effects could correlate with the independent variables. For instance, industries may have intrinsic characteristics like differing levels of unionization, regional economic factors, or industry-specific regulations that could simultaneously influence the extent to which they are subject to supply and demand shocks, and also impact wage and employment levels. If these unobserved effects do correlate with the regressors, the RE model can lead to biased and inconsistent estimates. In such cases, the FE model, which controls for time-invariant unobserved effects, would be the more appropriate choice as it eliminates potential bias arising from correlation between unobserved effects and regressors.

The key feature of the FD model is that it removes time-invariant unobserved individual (in this case, industry-specific) effects by differencing the data. This means that the model focuses on changes within an industry over time, rather than levels. This can be particularly useful in the context of your study on the effects of supply and demand shocks on hourly wages and employment. In the presence of supply and demand shocks, what typically matters more are the changes these shocks produce in wages and employment, not necessarily the absolute levels. By focusing on changes, the FD model directly captures the dynamics of how these shocks propagate through the economy. The FD model's ability to remove time-invariant unobserved effects can also be beneficial. Like the FE model, the FD model can control for any unobserved, time-invariant industry characteristics that may be correlated with the independent variables. This can help to mitigate the omitted variable bias that could arise if these characteristics were not properly accounted for. Moreover, the FD model is less sensitive to measurement errors compared to the FE model. This can be an advantage in situations where data accuracy is a concern. However, it's important to be aware of the potential limitations of the FD model. First, by differencing the data, the model may increase the noise-to-signal ratio, reducing the efficiency of the estimators. Second, the FD model removes the time-invariant effects entirely, making it unsuitable if these effects are of interest. Lastly, the FD model can exacerbate issues of autocorrelation and it also leads to the loss of one time period of data due to differencing.

The FD model eliminates time-invariant effects entirely, which can be both a strength and a limitation. If these effects are of interest, the FD model would not be appropriate. On the other hand, the FE model controls for time-invariant effects but retains their influence in the model, making it more suitable if these effects are an essential aspect of the analysis. The first differencing process can exacerbate issues with autocorrelation. If errors in the original data are already correlated over time, the first difference might increase this correlation, creating a new issue to address. FE model does not inherently exacerbate autocorrelation problems.

8 Conclusion

We can conclude that both demand and supply shocks exert influence on employment, although their impacts manifest in complex and sometimes counterintuitive ways.

Supply shocks, proxied by changes in production, have a significant positive effect on employment. When production increases, this leads to an upsurge in employment, as more labor is required to facilitate this increased production. This outcome aligns with traditional economic theory, suggesting that positive supply shocks (for instance, advancements in technology or an increase in availability of raw materials) that boost production, tend to also drive up employment. Firms need to hire more workers to keep up with increased production levels, hence expanding the job market within their respective industries. On the contrary, demand shocks, represented by Germany imports, display an inverse relationship with employment. Although we initially expected that an increase in demand from Germany (higher imports) would stimulate employment in the Danish industry, the empirical evidence suggests otherwise. An increase in Germany's demand for goods is associated with a decrease in employment in the industry under study.

Several possible explanations for this outcome were proposed. One possibility is that Danish firms are responding to increased demand from Germany by automating their production processes or utilizing more technology, which could reduce demand for labor. Alternatively, firms might relocate their production or certain stages of production to other countries where labor or other production costs are cheaper, in an effort to respond to the increased demand more efficiently. Lastly, increased demand might also be satisfied through the use of imported materials rather than domestic labor, leading to a decrease in employment.

The supply shock, proxied by changes in production, have a significant negative effect on hourly wages. When production increases, this surprisingly leads to a decrease in hourly wages. There could be several reasons for this counter-intuitive result. One possible explanation is that as firms increase production, they invest more in technology or equipment, reducing the dependence on labor and thereby exerting downward pressure on wages. Another possibility is that firms are achieving economies of scale as they increase production, and this increased efficiency may lead to a decrease in average wages.

Due to insignificance in the Germany imports variable, we are unable to conclude anything about the impact of demand shocks on hourly wages in this analysis. In economic theory, increased demand (demand shock) is usually expected to lead to an increase in both employment and wages, as firms need to attract more workers to meet the increased demand. However, since we do not have significant results for the Germany imports variable in this model, we cannot make any valid inferences about the effect of demand shocks on hourly wages.

However, we do not have sufficient information to definitively conclude why this negative relationship between production (our proxy for supply shocks) and hourly wages exists based on this data alone. Further investigation into the exact nature of these production increases and their impacts on the labor force would be needed to fully understand this phenomenon.

In conclusion to the analysis of the impact of demand and supply shocks on both employment and hourly wages revealed intriguing insights. It was evident that supply shocks, as proxied by changes in production, showed diverging effects on employment and wages. An increase in production was associated with an increase in employment, reflecting the direct relationship between production and labor demand. However, surprisingly, it was associated with a decrease in hourly wages, suggesting possible efficiency gains or increased technology use in the production process.

As for the demand shocks, represented by Germany imports, we observed that an increase in demand led to a decrease in employment in the first model, possibly due to firms adapting to higher demand by automating or outsourcing production. Unfortunately, due to insignificance in the Germany imports variable in the second model, we were unable to draw any inferences about the impact of demand shocks on hourly wages.

Overall, these results underline the complex and nuanced interplay between demand and supply shocks and their implications on labor market indicators, specifically employment and wages. Each shock seems to create a distinct ripple effect in the labor market, influencing employment and wages in varied and sometimes unexpected ways.

9 Bibliography

1. Source (2023), Bilateral Trade in Goods by Industry and End-use (BTDIxE), https://stats.oecd.org/Index.aspx DataSetCode=BTDIXE_I4# (04-05-2023).

2. Source: Danmarks Statistik - https://statistikbanken.dk/ NABP117: 1-2.1.1 Produktion, BVT og indkomstdannelse (117-gruppering) efter transaktion, branche og prisenhed.

3. Source: Danmarks Statistik - https://statistikbanken.dk/ NIO1: Input-output tabel. Tilgang fra brancher fordelt på anvendelse og prisenhed.

4. Source: Danmarks Statistik - https://statistikbanken.dk/ NIO5: Beskæftigelse og timer efter socioøkonomisk status og branche.

5. Source: Jeffrey M. Wooldridge (2018). Introductory Econometrics: A Modern Approach (7e international student ed.), Student textbook, Cengage Learning 2018.

6. Source: Hal R. Varian (2019). Intermediate Microeconomics with Calculus: A Modern Approach, Student textbook, WW Norton & Co

7. Source: Eric J. Pentecost (2000). Macroeconomics, Student textbook, Bloomsbury Publishing PLC