Essays in Applied Macroeconomics: A Regional Perspective Using US County-Level Data

Eduardo Teixeira de Carvalho Silva

A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy in Economics.
School of Economics, University of Kent

September 2025

Acknowledgements

I am sincerely grateful to the many people who supported me throughout this academic journey. First and foremost, I would like to thank my supervisor, Professor Miguel León-Ledesma, whose guidance and support over the years helped shape this dissertation from the very beginning. I am also grateful to Dr. Christian Siegel, who kindly agreed to supervise me during the later stages of my PhD. I would like to thank all the staff of the School of Economics for their ongoing support, with special thanks to Dr. María García-Alonso, Dr. Wei Jiang, and Dr. Sylvain Barde for the many encouraging and fruitful conversations over the last five years. I am also thankful to Dr. Alex Luiz Ferreira, who introduced me to Miguel and the School of Economics at the University of Kent years ago, opening the door that ultimately led to the completion of this PhD.

I am deeply grateful to the wonderful friends I made during my years at the University of Kent. Their support, kindness, and shared laughter made this journey not only more bearable during difficult moments but also genuinely joyful. I thank them for standing by me in the hard times and for celebrating each achievement along the way. I would especially like to thank Alexandra Kirtley, Chiara Zazzarino, Isabella Pellien, João Pedro Mainente, Maliha Rahanaz, Stuart Henderson, Zeynep Genc, and Zhan Shi for their friendship and the many memorable days we shared in the office. I also owe my deepest gratitude to my family and friends back in Brazil for their unwavering love and encouragement throughout this long journey.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In recent years, the empirical macroeconomics literature has increasingly turned to regional and other sorts of disaggregated data for empirical analysis as an alternative to using aggregate data. This dissertation comprises three essays that leverage the variation present in granular datasets to conduct macroeconomic analysis from a disaggregated perspective. More specifically, all essays use US county-level data for the majority of the analysis, exploiting its detailed variation in comparison to national-level data. In part of the analysis, the studies presented in this dissertation also resort to sectoral-level data.

This approach offers several advantages. Working at smaller geographic units yields greater cross-sectional and temporal variation, as well as many more observations, which helps mitigate challenges inherent in aggregate analysis, such as endogenous policy responses. Disaggregated data reveal heterogeneous effects and dimensions of economic behaviour that national aggregates mask. The micro-to-macro lens also enables explicit analysis of features such as commuting, migration, input–output networks, and firm composition, thereby clarifying the mechanisms behind aggregate outcomes.

The first essay of this dissertation, presented in **Chapter 2**, is named "*Estimating the US Wage Phillips Curve: a local labour market approach*". This essay estimates a wage Phillips curve using county-level labour market data. Its main innovation is the explicit modelling of spatial spillovers in local labour markets, allowing shocks in one county to affect wage inflation in neighbouring areas. The premise is that firms recruit across local labour markets rather than strictly within county borders, and workers similarly search beyond their county of residence, an assumption consistent with substantial inter-county commuting in the United States. Using county-level data helps address identification challenges arising from endogenous aggregate policy responses that lean against demand-side shocks. To mitigate remaining bias from cost-push disturbances, the wage Phillips curve is estimated with shift–share instruments constructed from sectoral labour market data. The analysis also incorporates the concepts of Commuting Zones and Labour Market Areas, using them to construct spatial neighbouring matrices. This approach shows that these definitions better capture local labour market integration than simple contiguity matrices. The chapter concludes by mapping the estimated county-level relationship into a national wage Phillips curve, highlighting important implications of modelling the spillovers into the aggregate curve.

The following essay, presented in **Chapter 3**, is titled "*A procurement-based dataset on the US military expenditure*". This essay documents and discusses a military expenditure dataset constructed from Department of Defence procurement records, highlighting its potential for applied macroeconomic research. Military spending is commonly used in fiscal policy analysis because it is arguably more exogenous than many other forms of public expenditure, being driven primarily by geopolitical considerations rather than local business cycles. Recent studies increasingly rely on procurement data given its high granularity, which enables detailed analysis of mechanisms that aggregate series obscure. The chapter makes two contributions. First, it provides a guide to the raw procurement data and a suite of R functions and codes to process this extensive dataset, including tools to aggregate data along geographical, sectoral, and firm dimensions, along with some additional analyses pertinent to typical applications of these data. Second, it offers a descriptive overview of aggregate procurement expenditure, such as its evolution over time (with links to political and economic developments), and its sectoral and geographical distribution, emphasising features relevant for empirical macroeconomic work.

Chapter 4 presents the third and final essay, titled "*The regional and sectoral labour market effects of military procurement expenditure*", which builds upon the data presented in the previous chapter, examining the impacts of military procurement expenditure on local labour markets. In the chapter, expenditure elasticities are estimated with respect to labour market outcomes such as employment and wages. Distinct dimensions and features of the regional markets are exploited, such as migration and intersectoral dynamics. The results indicate that higher procurement spending attracts workers to the affected areas, with in-migration helping to fill newly created vacancies. In addition, there is evidence of both spatial and cross-sectoral spillovers; however, these sectoral effects are heterogeneous across industries, highlighting the importance of production networks and local economic structure in shaping the transmission of procurement shocks.

Taken together, the essays in this thesis advance the empirical macro literature from a regional perspective, with an emphasis on labour markets, and introduce a procurement-based expenditure dataset with broad applicability for studies that exploit regional and sectoral granularity in analysing fiscal policy. The remainder of the thesis is organised as follows: Chapters 2–4 present the three essays, and Chapter 5 offers the concluding remarks.

Chapter 2

Estimating the US Wage Phillips Curve: a local labour market approach

2.1 Introduction

In this chapter, I estimate a wage Phillips Curve (WPC) for the United States using county-level labour market data, in a setting that accounts for the presence of geographical spillovers in the labour markets. The use of data at this disaggregation level overcomes problems commonly arising in the estimation with aggregate data, such as a lack of variability and an endogenous monetary policy bias. It also allows for an analysis of how local labour markets are interconnected and the dynamics between them. The empirical strategy incorporates shocks originating in neighbouring regions. Such spillovers are expected, given that workers commonly commute and search for jobs beyond their place of residence. I also show how the regional WPC with spillover effects maps into a nation-level curve. To my knowledge, those are novel contributions to this literature.

The relationship between wage inflation and unemployment plays a crucial role in shaping the dynamics between prices and economic activity, as it directly impacts the persistence of inflation. Wage inflation may perpetuate the impact of temporary macro shocks on inflation by exerting pressure on prices through the labour market channel. Thus, the response of wages to unemployment shocks remains a key building block of macroeconomic models for monetary and fiscal policy analysis.

The high unemployment generated by the US Fed's deflationary policies in the early 1980s was followed by a period of low and stable inflation. This Great Moderation period raised doubts about the stability, and even existence, of the Phillips Curve (PC) and has been the subject of intense debate, a debate that also extends to the wage Phillips curve. The apparent disappearance of the PC's can be attributed to the challenges that the use of aggregate data imposes on the identification of the relevant parameters of the curve.

Recent influential empirical works have examined the issues associated with using aggregate data in the estimation of Phillips curves. McLeay and Tenreyro (2020) and Fitzgerald et al. (2024) indicate that an endogenous monetary policy is responsible for the identification problem. As the Phillips Curve represents a supply curve, it requires demand-side shocks for its identification. Because the observed data represents equilibrium, one might fail to identify the true slope of the curve, as cost-push shocks will bias the estimation of the Phillips Curve coefficient towards zero. This identification problem is worsened under an optimal monetary policy setting, as there is no trade-off between stabilising output and inflation when a demand shock hits the labour market, so the monetary authority should always choose to counter demand-side shocks. That leaves only supply-side shocks to affect prices, which may confound the identification problem even further, possibly resulting in the estimation of a near zero, or even positive, relationship between unemployment and wages/prices.

As previously mentioned, limited variability is also indicated in the literature as a responsible for the identification problem, as aggregate time series data often lacks sufficient observations and variability for precise estimation. For instance, Hooper et al. (2020) note the limited variation in aggregate labour market data, as there are very few instances where the national unemployment rate significantly falls below the natural rate following the Volcker era. The expanding body of literature on this subject has addressed these issues by using disaggregated data. It contains the necessary granularity to capture important labour market characteristics, which is usually lacking in macro-level datasets. Regional data can reflect the heterogeneity that national-level data might overlook.

McLeay and Tenreyro (2020) emphasize that using local labour market data is particularly valuable for estimating Phillips curves since monetary policy responds to aggregate shocks, while local shocks can be considered exogenous and effectively used to identify a PC. Leveraging regional-level shocks allows estimating a PC free from biases associated with endogenous policy responses.

Hazell et al. (2022) pointed out another benefit of utilizing regional data: it allows for the use of a panel setting, which can incorporate fixed effects that account for shifts in long-run expectations, which can confound the estimation of a Phillips Curve. Those fixed effects also control for other unobservable variables, such as regional natural unemployment rate variations. As discussed by Hooper et al. (2020), including year and geographical fixed effects differences out aggregate time series variations and cross-sectional differences associated with unobservable variables. This implies that the variation used to identify the regional Phillips Curve differs from the variation used to identify the national-level curve.

The available data in the American labour markets is becoming increasingly detailed and disaggregated, enabling analyses at highly granular levels. Recent studies adopting this approach have used regional US data across various geographical scales, including state, city, and metropolitan statistical areas (MSA). However, to my knowledge, there has been no research that used data at the county level. While price data is not accessible at this disaggregation level, wage and employment datasets are available for the American counties. I draw on a highly detailed employment dataset from the Quarterly Census of Employment and Wages provided by the Bureau of Labor Statistics, encompassing 3,101 counties from 1991 to 2021.

County-level data offers a broader geographical perspective compared to city and MSA data, as it covers the area of the entire country. An important consideration when using county-level labour market data is that workers frequently commute or even relocate between counties for work. Local labour market pressure can impact the wage level of nearby areas. For instance, if the labour market conditions are tighter in a county, local companies may compete for workers from adjacent counties, thereby creating upward pressure on wages not only within that county but also in surrounding areas. This implies that the effects of labour market shocks are not limited to the local market but may spillover to neighbouring counties. Consequently, shocks affecting neighbouring locations might be relevant in explaining local labour market outcomes.

Based on the U.S. Census Journey to Work data, approximately 28% of American workers in the last two decades have commuted to work in other counties[1]. Additionally, people also migrate between counties. From 2005 to 2022, around 4% of the American population migrated across counties yearly[2], primarily searching for better job opportu-

[1] This approximation was calculated using county-to-county commuting flow data from the U.S. Census (`https://www.census.gov/topics/employment/commuting/guidance/flows.html`). It represents the proportion of workers from each county who commute for work. A proportion of around 28% was observed when analysing data from various time periods: 2006-2010, 2009-2013, 2011-2015, and 2016-2020.

[2] The source of this information is `https://shorturl.at/a84yk`, which uses data from the U.S. Census Bureau Current Population Survey.

nities or responding to declining labour market conditions in their home counties.

Using county-level data to estimate a wage Phillips Curve would then introduce a potential econometric concern. As workers commute and migrate between counties, spatial correlation might exist in county-level labour market data. If the neighbouring labour market conditions help explain local wage inflation, we might have a case of an omitted variable, violating classical Ordinary Least Squares assumptions and resulting in biased estimations. To handle this potential issue, I estimate an empirical WPC specification that allows local wage inflation to be affected by the unemployment rate from neighbouring counties. Additionally, I perform estimations using shift-share instruments to address any remaining local-level cost-push shocks that persist despite using disaggregated data.

Furthermore, I delve into the concepts of Commuting Zones (CMZ) and labour market areas (LMA) as defined by the Economic Research Service (ERS) to analyse the spatial interactions among local labour markets. Initially, I demonstrate that when data is aggregated at the labour market area level, the unemployment rates of neighbouring LMAs do not explain local wages[3]. However, since the coefficient for neighbouring unemployment rates is significant in regressions at the county level, this finding implies that any labour market spillovers are likely confined within the LMAs, making them suitable candidates for establishing a neighbouring matrix. Consequently, when estimating the regression at the county level, I account for all counties that fall within the same LMA as neighbours.

As Beraja et al. (2019) highlight, aggregating regional models to the macro-level is not trivial, as it may overlook relevant channels and connections. I propose an aggregation of the regional WPC to a national-level curve that accounts for the spillover effects between local labour markets. The aggregation suggests that not only should the size of a county's labour force be considered in its weighting for the overall aggregation, but also a measure of centrality. Counties with more neighbours hold greater relative importance, and their significance increases when their neighbours have fewer connections, meaning that shocks from a limited number or even no other regions can impact those neighbours. This finding carries significant policy implications, as the labour market dynamics in certain counties

[3] As demonstrated later, the same result does not hold for the aggregation at the CMZ level.

may exert a more substantial influence on the national-level wage inflation.

The estimation results for the national-level WPC coefficient on unemployment range from -0.292 to -0.325 with an Ordinary Least Squares estimation and from -1.063 to -1.657 when a two-stage estimation with shift-share instruments is used. These results exhibit, on average, greater magnitude compared to other studies in the literature that estimate a wage-price curve (WPC) using U.S. regional labour market data. This discrepancy can be attributed to the inclusion of spatial spillovers in the empirical model, which increases the aggregate coefficient of the WPC. Additionally, the application of Bartik-style instruments further increases the magnitude of the estimated coefficients.

This study builds upon existing literature that employs regional variation to estimate wage and price Phillips Curves. Different studies have used regional data from the United States at various levels of aggregation and through different methodologies to estimate wage Phillips Curves. Leduc et al. (2017) use US city-level data to examine if there were any changes in the slope of the WPC over the last decades, finding evidence of a flattening curve after 2009. Smith (2014) uses state-level data to estimate a WPC using different measures of slack constructed from labour market data including unemployment rates as in this study, but also other measurements such as short and long-term unemployment. Burya et al. (2023) estimate a wage PC for the United States using labour market data at the commuting zone level derived from a highly disaggregated dataset of online vacancy postings. They find that the WPC is flatter in commuting zones where firms have a larger labour market power over job seekers.

Other studies, such as Fitzgerald et al. (2024) and Beraja et al. (2019), imply the WPC from the parameters of a New Keynesian model. The authors employ state-level data to conduct a Bayesian estimation of Calvo wage and price parameters. Beraja et al. (2019) findings indicate that wages exhibit greater flexibility when regional data is integrated into the estimation process instead of relying solely on aggregate data.

Another branch of the literature argues that the WPC is non-linear, as wages may exhibit downward rigidity, and consequently, wage inflation may not respond symmetrically to unemployment rates exceeding or falling below the natural rate. Given its increased variability, regional data can be used to uncover non-linearities in the Phillips curve. For instance, Kumar and Orrenius (2016) and Hooper et al. (2020) analyse state-level data and find evidence supporting a non-linear WPC. Additionally, Hooper et al. (2020) and Babb and Detmeister (2017) identify non-linearities in the price Phillips Curve using Metropolitan Statistical Area (MSA)-level data, while similar analysis at the national level does not yield the same results. Furthermore, Bishop et al. (2021) discover non-linearities in the WPC through their research using Australian local labour market data.

The remainder of this chapter is organized as follows: Section 2 describes the dataset used in this study and provides a descriptive analysis of the data. Section 3 discusses the empirical model, presents the results and proposes an aggregation of the county-level WPC with spatial spillovers to a national-level curve. It also presents a sample splitting exercise that tests the stability of the WPC across time. Section 4 concludes the chapter.

2.2 Data and descriptive statistics

This section introduces the data used in this chapter and highlights some features pertinent to the empirical analysis. I first describe the data sources and provide summary statistics. Then, I use the data at different levels of aggregation to exemplify the identification problems associated with using national-level data. To finish the section, I bring a visual analysis of the spatial distribution of the data and spatial correlation tests to motivate the inclusion of spatial spillovers in the empirical model.

2.2.1 Data

Wages. The data originates from the Quarterly Census of Employment and Wages (QCEW), a program administered by the Bureau of Labor Statistics (BLS) that publishes statistics on employment and wages. These statistics are presented at the industry level, classified according to the 6-digit NAICS, and encompass over 95% of the total jobs in the United States. Data is available at various geographical levels, including national, state, county, and metropolitan statistical areas (MSAs).

The QCEW microdata is derived from administrative records from the state unemployment insurance (UI) program. All private and public sector employers under the UI are required to submit a Quarterly Contribution Report (QCR), which serves as the primary data source for the QCEW. Additionally, the program conducts two surveys: the Annual Refiling Survey (ARS) for firms with more than three employees and the Multiple Worksite Report (MWR) for businesses operating multiple locations. These surveys yield supplementary information, providing accurate details about the physical locations and economic activities of firms participating in the UI programs[4].

The QCEW covers a substantial number of enterprises and jobs throughout the United States. In 2019, it covered 10.23 million establishments, accounting for approximately 148.1 million jobs. The Bureau of Labor Statistics (BLS) employs a technique known as imputation to manage data from non-responding establishments. When data for a particular month is missing, the BLS uses the data from the previous month, adjusted by its growth rate from the prior year. Notably, the number of imputed values does not appear to be of significant concern. From the first quarter of 2001 to the fourth quarter of 2019, the percentage of imputed wages was small, varying between 1.77% and 4.78%.

The measurement of wages I use is the *annual average pay*. According to the QCEW, wage statistics are reported quarterly and reflect the total compensation received by workers during a quarter, irrespective of when the work was performed. This compensation includes all forms of remuneration, including tips, bonuses, the cash value of meals, and profit distributions. One vulnerability of this wage measure is that it is not reported on an hourly basis. As a result, fluctuations in this variable may stem from changes in the number of hours worked, rather than from changes in wages alone. It is also worth noting that this data refers to the county where the workplace is located, rather than where the worker lives. This means that this measure of wages captures the amount being paid by the local firms, regardless of where the worker resides.

To demonstrate that the QCEW's *annual average pay* serves as a reliable proxy for

[4] Detailed information regarding the construction of QCEW data can be accessed at `https://www.bls.gov/opub/hom/cew/home.htm`.

wage rates, I calculate the correlation between the wage inflation data from the QCEW and the growth rate of *average hourly earnings* data from the Current Employment Statistics (CES) survey. Although the latter is not available at the county level, it can be accessed at both the national and state levels. Prior to computing the correlations, I apply a Baxter-King (BK) filter to the data. The resulting correlation levels are notably high, measuring 0.651 at the state level and 0.942 at the national level. This suggests that a significant portion of the variation in QCEW earnings data is driven by fluctuations in wage levels rather than in the number of hours worked.

Labour Force and Unemployment. While the QCEW provides data on employment and wages, it does not offer metrics for the labour force and unemployment, which prevents the calculation of unemployment rates. Therefore, I rely on data from the Local Area Unemployment Statistics (LAUS).

Since LAUS does not conduct surveys, its data is derived from monthly estimates of employment and unemployment provided by the BLS. The methodology employed is known as the "Handbook method." This approach uses various available data sources, including the QCEW, CES, Current Population Survey (CPS), and population estimates from the U.S. Census Bureau, to estimate labour force, employment, and unemployment metrics for nearly 7,500 subnational areas. The Handbook method independently estimates each category of employed and unemployed individuals, subsequently combining these figures to generate total estimates and the overall labour force level[5].

In contrast to the QCEW data, the LAUS data is based on where people live, rather than the location where they work. Therefore, the unemployment rates provided represent the share of the labour force that resides in a given county and is actively searching for jobs. This allows for the possibility that an individual may be looking for work in neighbouring areas to the county where they reside, or even in more distant parts of the country.

To assess the reliability of the LAUS estimates, I conducted an analysis akin to that

[5]The methodology to estimate the Local Area Unemployment Statistics (LAUS) is explained in detail at `https://www.bls.gov/opub/hom/lau/home.htm`.

used for wage data. I calculated the correlation between the growth rates of employment from the LAUS and QCEW databases. Prior to computing the correlations, I applied a BK-filter to both sets of growth rates. This procedure was performed at each level of aggregation. The resulting correlation coefficients were 0.976 at the national level, 0.899 at the state level, and 0.671 at the county level. The smaller correlation for the county-level data is expected due to the previously mentioned characteristics of the QCEW and LAUS data: the first refers to the place of work, and the second to the place of residence.

Commuting Zones and Labour Market Areas. Data for CMZs and LMAs were produced by the Economic Research Service (ERS)[6]. The purpose of this dataset is to group counties into zones that reflect local economies and labour markets. The estimation of CMZs and LMAs follows the methodology introduced by Tolbert and Sizer (1996), using journey-to-work data from the US Census. The first delineation of CMZs and LMAs was based on data from the 1980 Census, with updates occurring every decade as new data becomes available.

The key distinction between CMZs and LMAs lies in their population size. Unlike CMZs, which are estimated without consideration of total population, LMAs are categorized based on a minimum population of 100,000 inhabitants. LMAs were considered by the ERS to be overly large and less effective than CMZs for capturing local integration, leading to their discontinuation after the 1990 Census.

The procedure to estimate the zones departs from the calculation of a metric of the connection between two counties, named proportional flow (P_{ij}):

$$P_{ij} = P_{ji} = \frac{f_{ij} + f_{ji}}{min\{L_i, L_j\}}. \tag{2.1}$$

In which f_{ij} and f_{ji} are the flow of commuters from county i to county j and vice versa. L_i and L_j denote the resident labour force of these two counties. A higher level of P_{ij} indicates a stronger commuting relationship[7]. With all values of P_{ij} established, a

[6]The data and descriptive information are available at the following links: `https://www.ers.usda.gov/data-products/commuting-zones-and-labor-market-areas/`; `https://sites.psu.edu/psucz/data/`

[7]If P_{ij} assumes a value greater than 1 it is decreased to 0.999, which indicates maximum connectivity.

proportional flow matrix can be computed. When $i = j$, P_{ij} is assigned a value of zero.

The proportional flow matrix is then transformed into a matrix of distance coefficients, referred to as dissimilarity matrix. The elements of this matrix are computed by subtracting each value from one, and all diagonal elements of the new matrix are set to zero:

$$\begin{cases} D_{ij} = D_{ji} = (1 - P_{ij}), & \text{if} \quad i \neq j \\ D_{ij} = 0, & \text{if} \quad i = j \end{cases} \tag{2.2}$$

CMZs are then delineated by clustering counties given their distance values from the dissimilarity matrix. Normally, a cut-off distance is predetermined. For the estimation of LMAs, the process is extended until all clusters contain at least 100,000 inhabitants.

Foote et al. (2021) highlight that the clustering of counties into CMZs and LMAs is sensitive to factors such as the selected cut-off distances[8] and measurement errors in commuting flow data. These factors can have implications for empirical results for studies that use CMZs and LMAs to define local labour markets, as demonstrated by the authors. With those considerations in mind, I use all available CMZ and LMA data in the empirical exercise as a robustness check. The following table summarizes this data:

Table 2.1: Commuting Zones and Labour Market Areas: summary statistics

Clustering Level	**CMZ**					**LMA**	
Year	*1980*	*1990*	*2000*	*2010*	*2020*	*1980*	*1990*
Number of zones	764	726	693	610	563	381	393
Average counties per zone	4.06	4.27	4.75	5.08	5.49	8.14	7.89
Median counties per zone	4	4	4	5	5	7	7
Max counties in a zone	22	19	19	20	17	34	35
Number of singletons	90	56	39	31	32	3	3

Notes: *The table presents summary statistics for differing county clustering levels, which were computed by the ERS using the US Census journey-to-work data from different years.*

Notice from the table above that LMAs, on average, contain a notably higher number of counties per zone than CMZs, leading to a smaller total count of zones. Moreover, the

[8] The chosen cut-off distance indirectly determines the total number of clusters.

overall quantity of zones in the CMZ dataset has diminished over time, resulting in larger zone sizes for the most recent data. Furthermore, the commuting zones data comprises many zones consisting of only one county, referred to as *singletons*.

Singletons present a challenge for the empirical approach of this chapter. As CMZs and LMAs are used to compute neighbouring matrices, the calculation of spatial lags becomes impossible for singletons, as these counties have no neighbouring entities. To address this issue, those counties were assigned to the CMZ or LMA with which they share the largest physical border.

Notice also that one downside of using CMZs and LMAs to group counties is that each county can only belong to a single zone. This arrangement prevents a county from being associated with another county that falls into a different group, even if the two share a geographical border. Consequently, LMAs might be preferable for some empirical analysis compared to CMZs since they typically encompass a greater number of counties and are less likely to overlook important labour market interdependencies.

2.2.2 Summary statistics of the data

The labour market data used in this chapter is at the yearly frequency, covering 3101 counties and ranging from 1991 to 2021, totalling 96131 observations. Counties with any missing values and those not found in both the LAUS and QCEW datasets were excluded from the analysis. Table 2.2 presents the summary statistics of the county-level wage inflation and unemployment rate data described previously. Wage inflation was calculated as the first difference of the logarithm of QCEW's annual average pay.

Table 2.2: County-level unemployment rate and wage inflation: summary statistics

Variable	*Mean*	*Median*	*St.dev*	*Max*	*Min*
Unemployment Rate	0.060	0.054	0.028	0.382	0.007
Wage Inflation	0.033	0.031	0.041	4.956	-0.823

Notes: *The table presents summary statistics for the unemployment rate and wage inflation measurements used in this chapter. The data is annual and at the county level, covering 3101 counties and ranging from 1991 to 2021, resulting in 96131 observations for each variable. The wage inflation denotes the first difference of the logarithm of the wage level.*

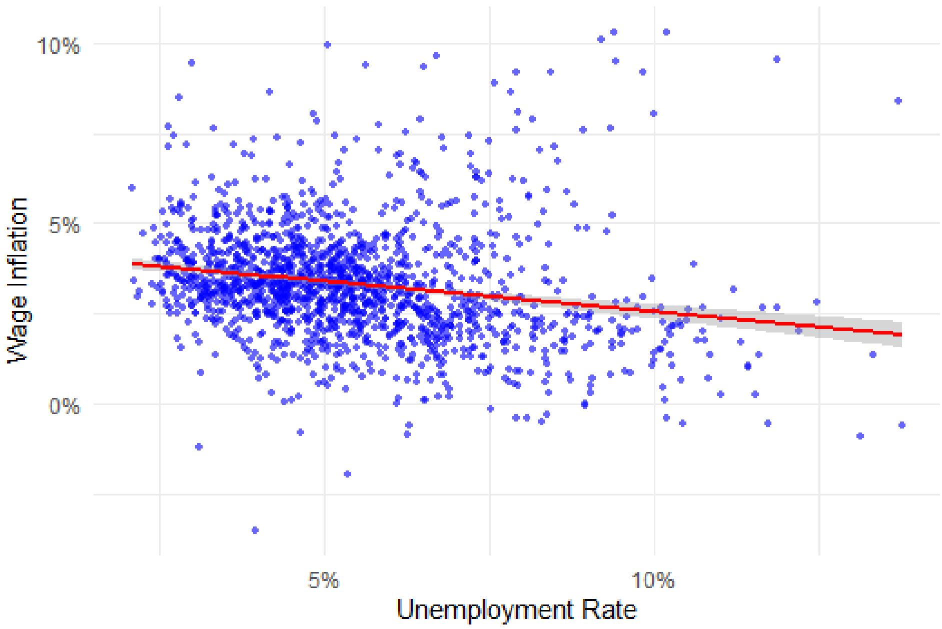
2.2.3 Illustrating the identification problem with aggregated data

The identification problem associated with using aggregated data is explored in this subsection. I aggregate the previously introduced county-level data at both the state and national levels. The following figures illustrate a scatterplot that depicts the relationship between wages and the unemployment rate at different levels of aggregation: national, state and county. These figures include a linear regression line with a shaded area around it representing a 95% confidence interval.

Figure 2.1: Wage inflation and unemployment: nation-level

Notes: *The figure shows a scatterplot of wage inflation versus the unemployment rate at the national level. The red line indicates a linear regression line for the data, while the shaded grey area represents a 95% confidence interval. The data ranges from 1991 to 2021, totalling 31 observations.*

Figure 2.2: Wage inflation and unemployment: state-level



Notes: *The figure shows a scatterplot of wage inflation versus the unemployment rate at the state level. The red line indicates a linear regression line for the data, while the shaded grey area represents a 95% confidence interval. The data ranges from 1991 to 2021, for 50 states, totalling 1550 observations.*

Figure 2.3: Wage inflation and unemployment: county-level

Notes: *The figure shows a scatterplot of wage inflation versus the unemployment rate at the county level. The red line indicates a linear regression line for the data, while the shaded grey area represents a 95% confidence interval. Observations of wage inflation exceeding 20% in absolute terms have been excluded from the figure. The complete dataset ranges from 1991 to 2021, for 3101 counties, totalling 96131 observations.*

Figures 2.1 - 2.3 suggest that the negative relationship between wage inflation and unemployment rates becomes more evident as the data is further disaggregated. The negative correlation is not statistically significant at the national level but is also stronger when we move from the national to the state and county levels. Apart from the clear gains in terms of the number of observations, there is also the limited variability in the national-level data, which lacks any episodes of negative wage inflation. These, however, get more frequent with higher data granularity.

To further support these features, I also estimate a simplistic set of WPC specifications at each aggregation level. The estimated equations are described below:

$$\Delta w_t = \alpha + \beta \Delta w_{i,t-1} + \gamma \left(U_t - U_t^N \right) + \epsilon_t \tag{2.3}$$

$$\Delta w_{i,t} = \alpha_i + \delta_t + \beta \Delta w_{i,t-1} + \gamma U_{i,t} + \epsilon_{i,t} \tag{2.4}$$

Equation (2.3) was estimated at the national level and (2.4) at the state and county levels. The discrepancy between specifications is because a measurement of natural unemployment rate is not available at the state and county levels. Δw denotes the first difference of the log of wages (wage inflation), and U is the unemployment rate. Despite the absence of a local measure of natural unemployment at these aggregation levels, the panel structure of (2.4) enables the incorporation of time (δ_t) and regional (α_i) fixed effects, thus accounting for variations in the unobserved natural unemployment rate. The measurement of the national natural unemployment rate (U_t^N) was sourced from the Congressional Budget Office[9]. The following table presents the results of the estimations:

[9] Data is available at `https://www.cbo.gov/data/budget-economic-data`.

Table 2.3: Equations (2.3) and (2.4) estimation results

Aggregation Level	***National***	***State***	***County***
α	0.023**		
	(0.007)		
β	0.339*	0.019	-0.127***
	(0.193)	(0.027)	(0.003)
γ	-0.092	-0.301***	-0.265***
	(0.193)	(0.035)	(0.010)
Number of observations	31	1550	96131

Notes: Estimation results of equations (2.3) and (2.4) for different levels of data geographical aggregation, which are indicated in the first row. The sample ranges from 1991 to 2021. The number of states in the dataset is 50, while the number of counties is 3101.
"***", "**" and "*" denote statistical significance at the 1, 5 and 10 per cent levels.

As the table above indicates, the WPC coefficient on the unemployment rate is smaller and not statistically significant when national level data is used. These results provide supporting evidence of the endogenous monetary policy described above, as it biases the estimated coefficient towards zero. Additionally, the precision of the estimates increases drastically with higher granularity, which is also a consequence of the larger number of observations available.

2.2.4 Spatial dependence in the data

In this subsection, I examine the presence of spatial dependence in the county-level labour market data. First, I present a visual analysis of the data, and then the results for a Moran's I test for spatial correlation.

Visual analysis of the data

This section illustrates the spatial distribution of the county-level unemployment rate and wage inflation in the USA[10]. Three different years were selected for the visual representation: 1999, 2009, and 2019. Those years were chosen because they are a decade apart and cover different labour market conditions. Specifically, 1999 and 2019 cover years of a

[10]The state of Alaska was not included in the analysis due to significant missing information from its counties.

heated market, while 2009 covers an economic downturn. Figures 2.4 - 2.6 illustrate the spatial distribution of the unemployment rates for the abovementioned years.

Figure 2.4: County-level unemployment rate (1999)

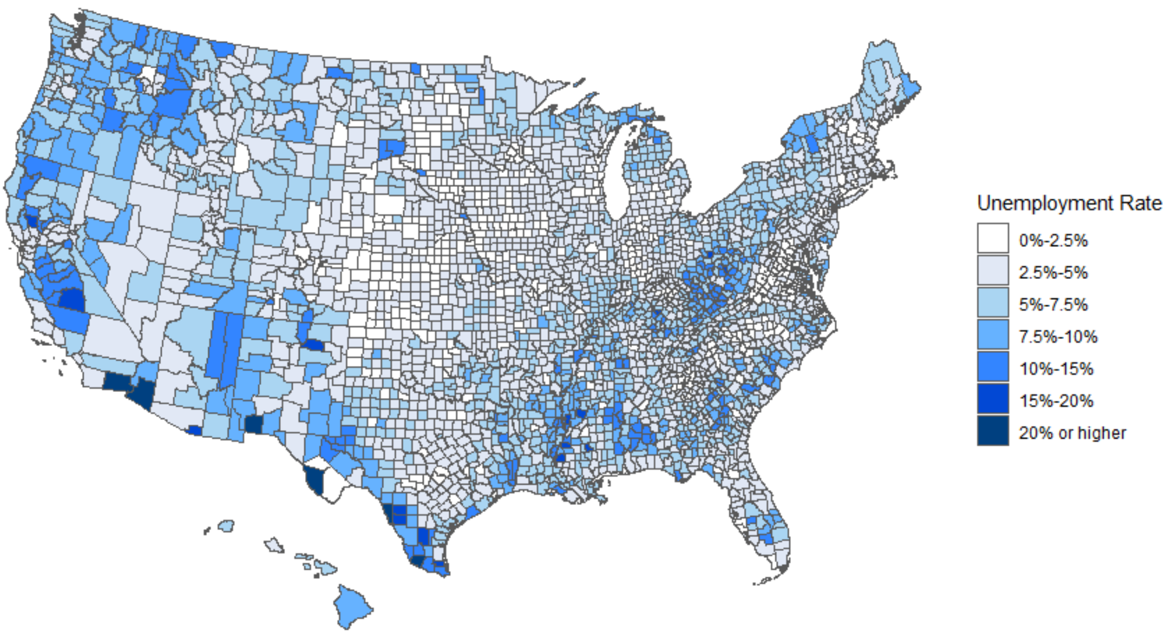


Figure 2.5: County-level unemployment rate (2009)

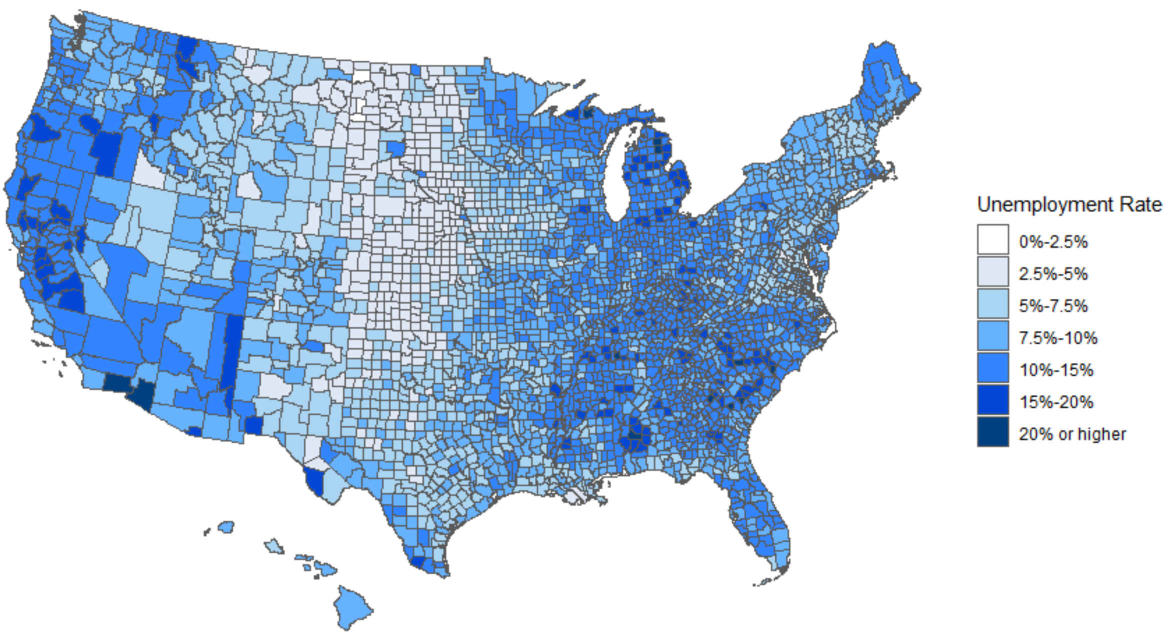
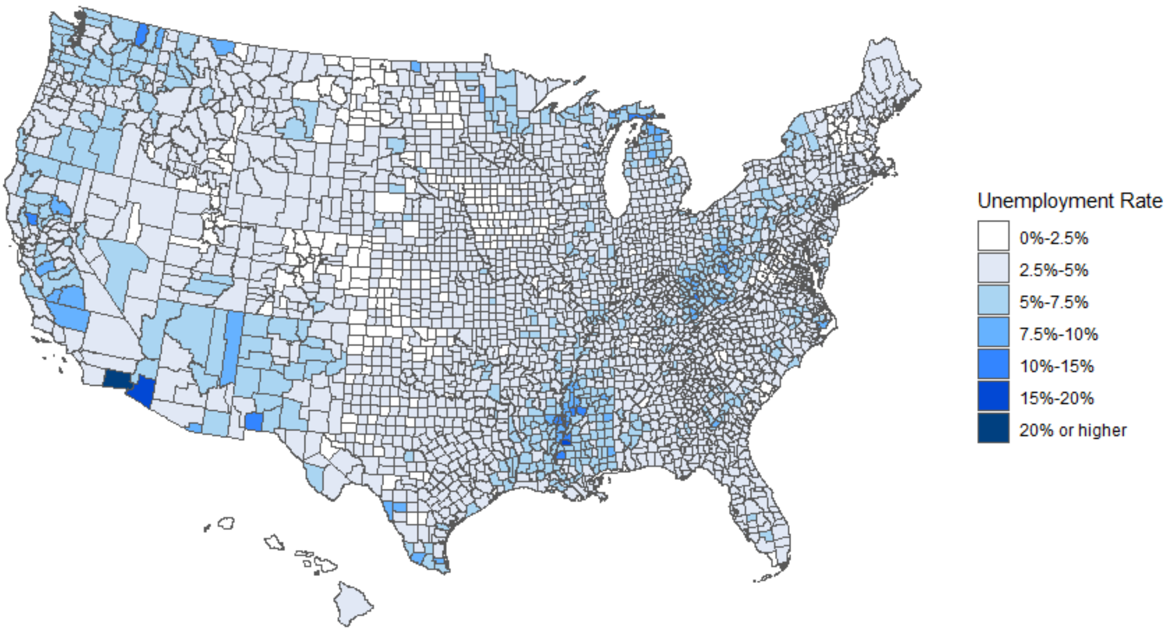


Figure 2.6: County-level unemployment rate (2019)



The maps above suggest a spatial dependence on the unemployment rate. For instance, the Midwestern region shows low unemployment rates in all three maps. While the clusters are less apparent in 2019, we can notice different clusters of high unemployment in the other two maps. For example, we can identify clusters in California and around Kentucky and West Virginia.

Figures 2.7 – 2.9 present the spatial distribution of county-level wage inflation for 1999, 2009, and 2019. Notice that the counties in yellow indicate those that faced wage deflation in that year.

Figure 2.7: County-level wage inflation (1999)

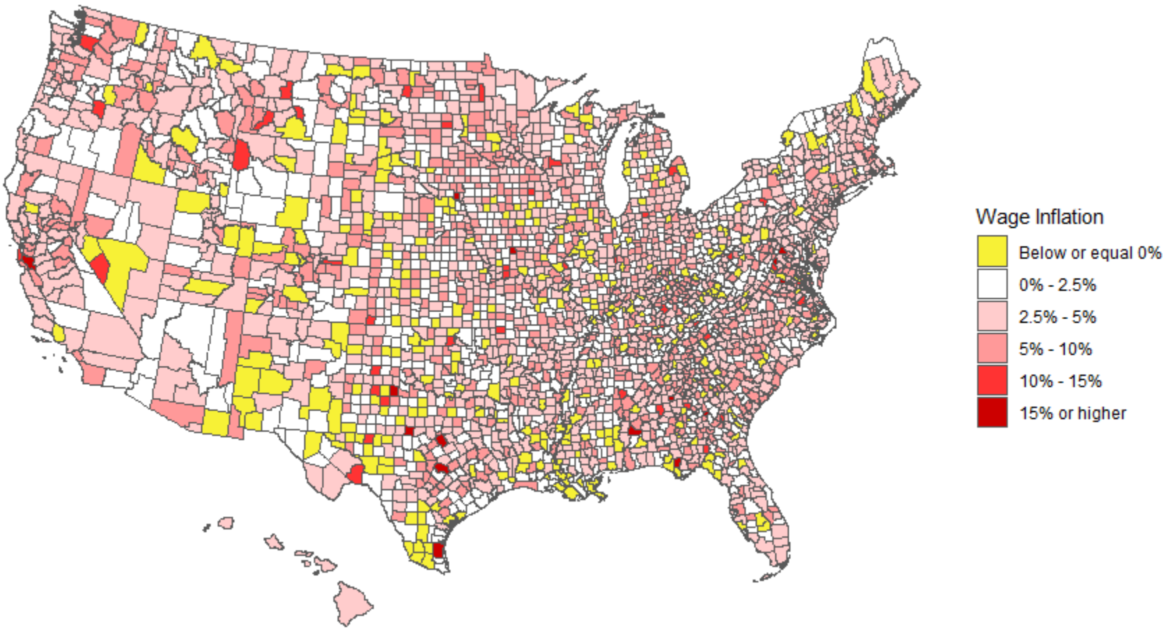


Figure 2.8: County-level wage inflation (2009)

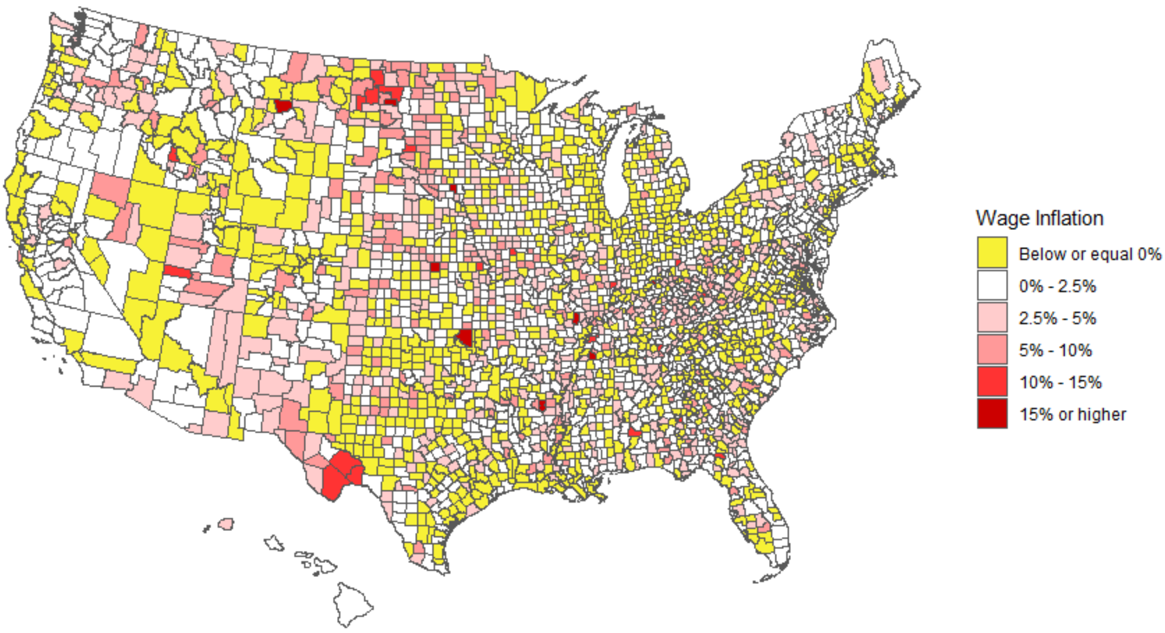
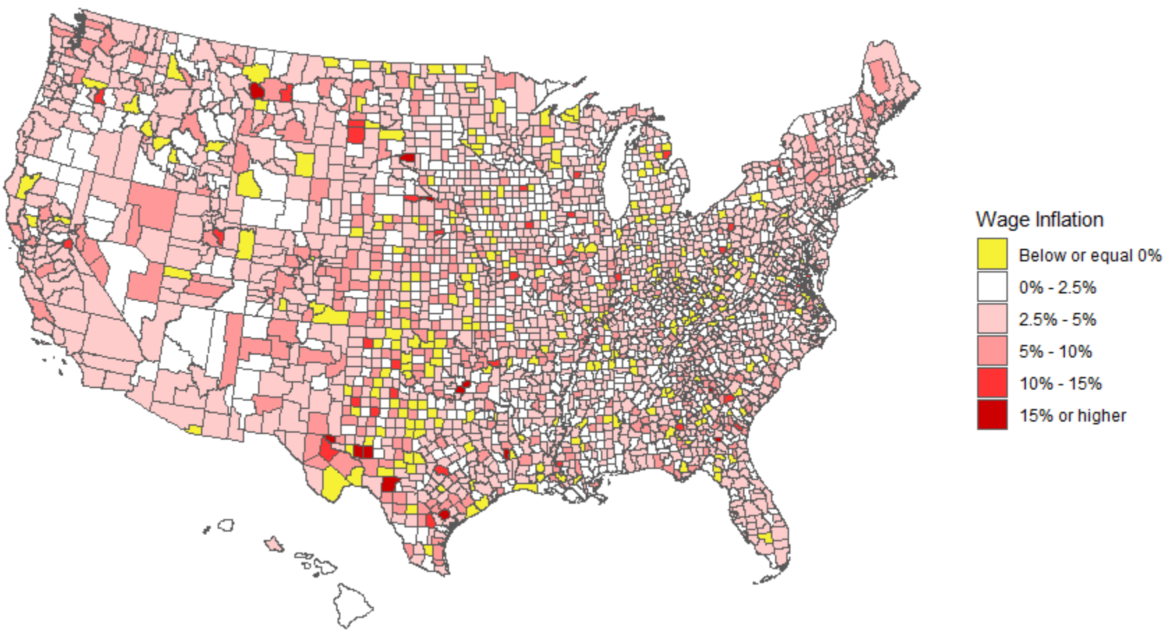


Figure 2.9: County-level wage inflation (2019)



The presence of clusters is less apparent for the wage inflation data. The year of 2019 appears to have more clusters, as we can identify more regions with neighbours facing wage deflation and also regions with groups of counties observing wage inflation of 10% or above. Also, during 2009 most the counties that showed wage deflation seem to be grouped together.

2.2.5 Moran's I test

In this section, I present a Moran's I test to examine the presence of spatial correlation in the data, complementing the analysis from the previous maps. The test was proposed by Moran (1950) and consists of a correlation coefficient between a variable in a given location and its neighbours. It was designed for a cross-sectional setting, taking the form:

$$I = \frac{X'WX}{X'X} \tag{2.5}$$

In which X is the variable of interest and W is a spatial weights matrix.

The following table presents the panel setting Moran's I test for the unemployment rate and for the wage inflation. W is defined as a contiguity matrix.

Table 2.4: Moran's I test

Year	***Unemployment rate***		***Wage inflation***	
	Test statistic	**p-value**	**Test statistic**	**p-value**
1991	0.547	0.001	0.011	0.117
1992	0.526	0.001	0.054	0.001
1993	0.524	0.001	0.020	0.011
1994	0.514	0.001	0.063	0.001
1995	0.493	0.001	0.000	0.412
1996	0.472	0.001	0.017	0.043
1997	0.484	0.001	0.011	0.066
1998	0.489	0.001	0.043	0.001
1999	0.489	0.001	0.034	0.004
2000	0.592	0.001	0.067	0.001
2001	0.571	0.001	0.039	0.001
2002	0.578	0.001	0.038	0.001
2003	0.571	0.001	0.046	0.001
2004	0.561	0.001	0.030	0.002
2005	0.605	0.001	0.113	0.001
2006	0.618	0.001	0.162	0.001
2007	0.625	0.001	0.117	0.001
2008	0.657	0.001	0.167	0.001
2009	0.694	0.001	0.112	0.001
2010	0.705	0.001	0.039	0.001
2011	0.699	0.001	0.206	0.001
2012	0.695	0.001	0.150	0.001
2013	0.665	0.001	0.020	0.025
2014	0.661	0.001	0.093	0.001
2015	0.633	0.001	0.134	0.001
2016	0.620	0.001	0.110	0.001
2017	0.602	0.001	0.049	0.001
2018	0.595	0.001	0.059	0.001
2019	0.596	0.001	0.043	0.001
2020	0.622	0.001	0.246	0.001
2021	0.627	0.001	0.053	0.001

Notes: *The table presents the cross-sectional Moran's I test results for spatial correlation for the county level unemployment rates and wage inflation for the years from 1991 to 2021. The spatial weights matrix used to compute the test statistic was a contiguity matrix.*

The test results indicates the presence of spatial correlation in the unemployment rates for all years and in the wage inflation for most years. However, the test statistic indicates a high spatial correlation in the unemployment rate, while a much weaker spatial correlation is found for the wage inflation. Given evidence of spatial correlation in unemployment, WPC regressions should control for neighbouring unemployment rates whenever they help explain local wage inflation; otherwise, the coefficient on local unemployment will be biased by omitted spatial dependence.

2.3 Estimation of the regional WPC

2.3.1 Methodology

I depart from the standard panel setting specification for the Wage Phillips Curve with time and region fixed effects as described by (2.4):

$$\Delta w_{i,t} = \alpha_i + \delta_t + \beta \Delta w_{i,t-1} + \gamma U_{i,t} + \epsilon_{i,t}. \tag{I}$$

Notice that the lag of wage inflation is being used to control for the wage inflation expectations, as there are no measures for it at the county level.

The panel structure of the specification provides a few advantages. The regional and time fixed effects, denoted by α_i and δ_t, respectively, control for permanent differences in inflation expectations across regions and for trends that are common to all counties over time. As highlighted in the previous section, the inclusion of fixed effects also controls for changes in the natural rate of unemployment, a variable that is not available at the county level. Moreover, as the local unemployment data used are values estimated by the BLS, they are susceptible to measurement error. Those, however, are absorbed by the model fixed effects.

Controlling for spatial interactions. Given that the county-level data spans the whole country and that the previous sections motivates the presence of spatial correlation in the data, I extend the baseline WPC specification (I) by incorporating spatial elements. This approach accounts for the influence of neighbouring labour market conditions on local wages and helps to reduce any estimation bias that may arise from omitting these spatial factors if they are relevant.

There are three ways of introducing spatial dependence in (I): through the independent variable, the dependent variables, and the error term. When a spatial lag of the dependent variable is incorporated into the model, interpreting the WPC coefficient becomes more complex due to the resulting feedback loop. To simplify the interpretation of the coefficients and the aggregation to a national-level WPC, I introduce spatial dependence in the model by adding the spatial lag of $\Delta w_{i,t-1}$ and $U_{i,t}$ (the dependant variables), and also allowing for spatial correlation in the error term. Two new specifications are proposed, described by the equations below:

$$\Delta \boldsymbol{w}_t = \boldsymbol{\alpha} + \delta_t \boldsymbol{\iota}_N + \beta_1 \Delta \boldsymbol{w}_{t-1} + \beta_2 W \Delta \boldsymbol{w}_{t-1} + \gamma_1 \boldsymbol{U}_t + \gamma_2 W \boldsymbol{U}_t + \boldsymbol{\mu}_t \tag{II}$$

$$\begin{aligned} \Delta \boldsymbol{w}_t &= \boldsymbol{\alpha} + \delta_t \boldsymbol{\iota}_N + \beta_1 \Delta \boldsymbol{w}_{t-1} + \beta_2 W \Delta \boldsymbol{w}_{t-1} + \gamma_1 \boldsymbol{U}_t + \gamma_2 W \boldsymbol{U}_t + \boldsymbol{\mu}_t \\ \boldsymbol{\mu}_t &= \lambda W \boldsymbol{\mu}_t + \boldsymbol{\epsilon}_t. \end{aligned} \tag{III}$$

The variables in bold represent the stack of the values for all N regions in a given period t. ι_N represent a vector of ones with size N. W represents a (NXN) spatial weights matrix[11]. Its elements are given by w_{ij}, which determines the relationship between regions i and j. If $i = j$ the value is set to zero, as a county cannot be a neighbour of itself.

When W pre-multiplies a variable, its spatial lag is computed. This consists of taking the average variable across a region's neighbours. (II) allows for the neighbouring average of the dependent values to affect local wage inflation. (III) does the same and allows for spatial dependence in the non-observable factors determining local wage inflation[12].

Different strategies can be employed to define W. I first estimate the model using a contiguity matrix, one of the simplest specifications. It equals one if two regions share a border and zero otherwise. The matrix is then row-normalized such that each row sums to one. I later move to alternative specifications for W other than a contiguity matrix.

[11] Also referred to as a neighbouring matrix.

[12] I discuss in the appendix A the downsides of using this specification, specially in the presence of fixed effects.

Bartik-style instruments. Despite the use of disaggregated data helping with the identification problem, there is still the possibility that changes in a county's wage inflation result from cost-push shocks, biasing the estimated WPC coefficient towards zero. To address this potential issue, I also estimate the empirical specifications presented in this section by instrumentalizing the unemployment rate with a Bartik-style (shift-share) instrument.

The construction of the instruments follows Goldsmith-Pinkham et al. (2020). It involves the interaction of the employment share in each sector for every county in 1992—the initial year of the estimation data, which is time-invariant —and the growth rate of employment in each sector at the national level. This is described by the equation below:

$$B_{it} = \sum_k z_{i,k,1992} g_{k,t} \tag{2.6}$$

In which $B_{i,t}$ denotes the instrument, $z_{i,k,1992}$ the sectoral (k) employment shares in each county (i) in 1992, and $g_{k,t}$ the sectoral growth rate of employment at the national level. The instruments were computed using the employment data from the QCEW.

Bartik-style instruments exploit the differential exposure of counties to national-level shocks, with this exposure being quantified by the county's initial industry employment shares. As noted by Borusyak et al. (2025), this style of shift-share instruments isolates demand-side variation as it keeps the local industry shares constant, while g can be interpreted as specific labour demand shifts. By using national industry growth rates, the instrument captures variation in local labour market conditions that stem from national economic trends, rather than from local factors, introducing a set of common shifts that should be predictive of the local industry growth rates while only capturing demand variation.

This structure isolates the component of local outcomes driven by national demand shocks, avoiding confounding from supply-side or region-specific shocks. As long as the initial shares are not themselves endogenous to future shocks, the Bartik instrument provides a reliable source of exogenous shifts in labour demand. As z is fixed in advance, it is plausibly unrelated to contemporaneous local outcomes.

2.3.2 Estimation results

Contiguity-based spatial weights

I first present the estimation of specifications (I) - (III) using a contiguity matrix to determine W. Two counties are considered neighbours if they share any common boundary point[13]. Table 2.5 presents the estimation results. (I) and (II) were estimated with the usage of an ordinary least squares (OLS) estimator, while (III) was estimated with a maximum likelihood (ML) estimator.

Table 2.5: Estimation results: contiguity matrix

Specification	(I)	(II)	(III)
Δw_{t-1}	-0.106***	-0.135***	-0.135***
	(0.017)	(0.014)	(0.003)
$W * \Delta w_{t-1}$		0.128***	0.117***
		(0.012)	(0.007)
U_t	-0.264***	-0.154***	-0.156***
	(0.009)	(0.024)	(0.013)
$W * U_t$		-0.171***	-0.170***
		(0.026)	(0.019)
$W * \mu_t$			0.150***
			(0.005)

Notes: Estimation results of equations I, II, and III when W is specified as a contiguity matrix. The first two specification were estimated using the OLS, while the third was estimating with a ML estimator. The data range is 1992-2021. The standard errors were clustered at the county level.
"***", "**" and "*" denote statistical significance at the 1, 5 and 10 per cent levels.

Table 2.6 presents the two-stage Instrumental Variables (IV) estimations for equations (I) – (III) using Bartik-style instruments for the unemployment rate. Given that neighbouring unemployment rates may also be affected by their own local cost-push shocks, I use the spatial lag of the fitted values from the first stage regression in the second stage instead of $\boldsymbol{WU_t}$.

[13] Although the Hawaiian counties do not share physical borders because they are islands, I defined them as neighbouring regions.

Table 2.6: IV estimation results: contiguity matrix (Bartik-style instruments)

Second Stage			
Specification	(I)	(II)	(III)
Δw_{t-1}	-0.155***	-0.160***	-0.161***
	(0.016)	(0.016)	(0.004)
$W * \Delta w_{t-1}$		0.036**	0.022*
		(0.017)	(0.012)
U_t	-1.275***	-0.476***	-0.580***
	(0.102)	(0.155)	(0.085)
$W * U_t$		-1.181***	-1.124***
		(0.156)	(0.119)
$W * \mu_t$			0.160***
			(0.005)
First Stage			
Δw_{t-1}		-0.020***	
		(0.001)	
$W * \Delta w_{t-1}$		-0.089***	
		(0.002)	
Bartik		-0.255***	
		(0.007)	

Notes: Results for the two-stage estimation of equations I, II, and III when W is specified as a contiguity matrix. The top part of the table reports the results of the second stage of the IV procedure, while the bottom part reports the first stage. The first two specification were estimated using the OLS, while the third was estimating with a ML estimator. The data range is 1992-2021. The standard errors were clustered at the county level.
"****", "***" and "*" denote statistical significance at the 1, 5 and 10 per cent levels.

As shown by the tables above, introducing the spatial lag of the unemployment rate in the regression leads to a reduction in the estimated coefficient on U_t, as is evident when comparing specification (I) to (II) or (III). As the unemployment rate is positively correlated with its spatial lag, omitting the latter results in a positive bias in the magnitude of the coefficient on U_t. The findings suggest that unemployment rates in neighbouring markets play a significant role in explaining local wage inflation. Notably, as shown in

Table 2.6, the unemployment rate and its spatial lag coefficients exhibit a marked increase when shift-share instruments are employed.

In Appendix A, I estimate the models (II) and (III) using alternative specifications for W. I use different forms of a fixed radius matrix and an inverse distance matrix. The estimation results indicate that the size of th neighbouring regions should be relatively small; otherwise, they start to mimic the nation-level variables, which are already captured by the time-fixed effects.

Considering the results presented above and in Appendix A, I conclude that specification (II) is preferred over (III) for several reasons. First, the estimation of (III) is computationally more intensive due to its structure. Second, the estimated coefficients on U_t and $W * U_t$ remain relatively stable upon introducing spatial dependence in the error term. Third, including spatial dependence impacts the behaviour of the time-fixed effects, severely decreasing their magnitude and significance. The fixed effects seem to capture most of the unobservable factors, making the spatial dependence in the error term redundant. Additionally, as discussed in Appendix A, there is a collinearity issue between the time-fixed effects and the spatial element in the error term, and this issue exacerbates with larger neighbouring regions. For the reasons stated above, the following results reported in this chapter will adhere to specification (II).

CMZ and LMA level estimations

In this subsection, I examine the suitability of the county blocks represented by Commuting Zones (CMZs) and Labour Market Areas (LMAs) for calculating the neighboring matrix W. These areas are promising candidates since they are derived from commuting data and should therefore reflect local labour market interactions. I first aggregate the data at the CMZ and LMA levels and re-estimate the model. If these groupings effectively capture local labour market interactions, we should observe a significant decline in the coefficients of the spatial components or their approach towards zero. In essence, the spillover effects should diminish, as these spatial interactions are anticipated to be contained within the designated zones.

As discussed in the Data section, the CMZ and LMA clusters vary depending on the measurement error in commuting flow data and the predetermined cut-off distances used to delimit them, having potential implications for procedures that use this data. As Foote et al. (2021) suggested, I estimate the model using all available CMZ and LMA data as a robustness test. Table 2.7 displays the OLS estimation results, while table (2.8) presents the IV estimation results.

Table 2.7: Estimation results: CMZ and LMA aggregation (OLS)

Aggregation Level	**Commuting Zone**					**Labour Market Area**	
Year	*1980*	*1990*	*2000*	*2010*	*2020*	*1980*	*1990*
Δw_{t-1}	-0.098***	-0.070***	-0.078***	-0.087***	-0.072**	-0.093**	-0.110***
	(0.030)	(0.027)	(0.027)	(0.030)	(0.032)	(0.041)	(0.039)
$W * \Delta w_{t-1}$	0.180***	0.213***	0.198***	0.190***	0.202***	0.186***	0.169***
	(0.030)	(0.025)	(0.026)	(0.030)	(0.027)	(0.032)	(0.042)
U_t	-0.235***	-0.259***	-0.291***	-0.278***	-0.303***	-0.271***	-0.307***
	(0.032)	(0.034)	(0.034)	(0.032)	(0.037)	(0.035)	(0.034)
$W * U_t$	-0.078***	-0.033	-0.026	-0.029	-0.001	-0.013	0.004
	(0.035)	(0.031)	(0.032)	(0.032)	(0.034)	(0.032)	(0.035)

Notes: Results for the OLS estimation of specification II with the data aggregated at different Commuting Zone and Labour Market Area levels. The second row indicates the date of the commuting flow data used to compute each clustering data. The standard errors were clustered at the CMZ/LMA level.
"***", "**" and "*" denote statistical significance at the 1, 5 and 10 per cent levels.

Table 2.8: Estimation results: CMZ and LMA aggregation (Bartik-style instruments)

Aggregation Level	**Commuting Zone**					**Labour Market Area**	
Year	*1980*	*1990*	*2000*	*2010*	*2020*	*1980*	*1990*
				Second Stage			
Δw_{t-1}	-0.131***	-0.100***	-0.113***	-0.117***	-0.103***	-0.117**	-0.133***
	(0.033)	(0.032)	(0.032)	(0.035)	(0.038)	(0.049)	(0.048)
$W * \Delta w_{t-1}$	0.053	0.104***	0.073**	0.074*	0.093**	0.113***	0.091*
	(0.038)	(0.034)	(0.035)	(0.039)	(0.038)	(0.041)	(0.054)
U_t	-0.667***	-0.585***	-0.643***	-0.470***	-0.670***	-0.600***	-0.758***
	(0.143)	(0.134)	(0.137)	(0.145)	(0.149)	(0.173)	(0.148)
$W * U_t$	-0.595***	-0.330**	-0.478***	-0.563***	-0.279	-0.023	0.090
	(0.169)	(0.155)	(0.175)	(0.169)	(0.207)	(0.211)	(0.175)
				First Stage			
Δw_{t-1}	-0.036***	-0.051***	-0.050***	-0.054***	-0.052***	-0.061**	-0.050*
	(0.012)	(0.013)	(0.013)	(0.017)	(0.017)	(0.026)	(0.029)
$W * \Delta w_{t-1}$	-0.148***	-0.179***	-0.161***	-0.170***	-0.166***	-0.198***	-0.181***
	(0.014)	(0.015)	(0.016)	(0.020)	(0.017)	(0.024)	(0.029)
Bartik	-0.373***	-0.468***	-0.479***	-0.479***	-0.473***	-0.611***	-0.635***
	(0.046)	(0.029)	(0.031)	(0.035)	(0.063)	(0.051)	(0.052)

Notes: Results for the IV estimation of specification II with the data aggregated at different Commuting Zone and Labour Market Area levels. The second row indicates the date of the commuting flow data to compute each clustering data. The top part of the table reports the results of the second stage of the IV procedure, while the bottom part reports the first stage. The standard errors were clustered at the CMZ/LMA level.
"***", "**" and "*" denote statistical significance at the 1, 5 and 10 per cent levels.

Tables 2.7 and 2.8 indicate that when the labour market data is aggregated at the CMZ and LMA levels the magnitude of the coefficients on $W * U_t$ significantly diminishes, losing significance in nearly all instances for the OLS estimations and in the last three cases for the IV estimation. In contrast, the coefficient on U_t shows an increase in magnitude.

The results indicate that LMAs effectively capture the integration of local labour markets and the spillover effects resulting from commuting. It suggests that firms tend to compete for workers within the counties of their respective zones, implying that unemployment within those zones is sufficient to account for local wage inflation. In contrast, unemployment from other LMAs does not significantly contribute to explaining local wage inflation.

Notice that the only three blocks in which the coefficient on $W * U_t$ is not significant for both OLS and IV estimations are CMZ 2020, LMA 1980, and LMA 1990. For the latter two, the estimated coefficient is very close to zero in both methodologies. As indicated in Table 2.1, these clustering blocks represent the ones with the highest average number of counties per zone. This suggests that if the zones are too small, they may not effectively capture the local labour market interactions resulting from commuting among those counties.

The areas defined by LMA 1980, and LMA 1990 appear to be sufficiently large to capture the dynamics of the local labour market, making them strong candidates for establishing the neighbouring matrix W. Those options offer a more structured and accurate approach regarding spatial spillovers compared to a contiguity matrix.

I now proceed by re-estimating the model at the county-level, but using the zones delineated by LMA 1980, and LMA 1990 to determine W[14]. CMZ 2020 is also used for a comparison of the results. If two counties, i and j, are part of the same zone, then $w_{i,j}$ will equal 1; otherwise, it will be 0. The OLS and IV estimations results can be found in the following tables:

[14]The CMZ and LMA datasets, as illustrated in Table 2.1, comprise counties designated as singletons. These singletons present a challenge for the methodology employed, as they are isolated within their groups without neighbouring counties, making it impossible to compute spatial lags. To address this issue, all singleton counties were incorporated into the group that shares the longest physical border with them.

Table 2.9: OLS estimation results: CMZ and LMA neighbouring matrices

Neighbouring Matrix	**CMZ 2020**	**LMA 1980**	**LMA 1990**
Δw_{t-1}	-0.133***	-0.133***	-0.133***
	(0.014)	(0.014)	(0.014)
$W * \Delta w_{t-1}$	0.100***	0.310***	0.124***
	(0.012)	(0.012)	(0.017)
U_t	-0.183***	-0.195***	-0.192***
	(0.023)	(0.023)	(0.023)
$W * U_t$	-0.126***	-0.097**	-0.109***
	(0.021)	(0.022)	(0.023)

Notes: OLS estimation results of (II) when the clustering data CMZ 2020, LMA 1980 and LMA 1990 are used to specify the neighbouring matrix W. The data range is 1992-2021. The standard errors were clustered at the county level.
"***", "**" and "*" denote statistical significance at the 1, 5 and 10 per cent levels.

Table 2.10: IV estimation results: CMZ and LMA neighbouring matrices (Bartik-style instruments)

Neighbouring Matrix	**CMZ 2020**	**LMA 1980**	**LMA 1990**
Second Stage			
Δw_{t-1}	-0.153***	-0.150***	-0.149***
	(0.016)	(0.016)	(0.015)
$W * \Delta w_{t-1}$	0.027	0.034	0.049
	(0.018)	(0.024)	(0.037)
U_t	-0.678***	-0.743***	-0.768***
	(0.151)	(0.163)	(0.175)
$W * U_t$	-0.614***	-0.406**	-0.295
	(0.158)	(0.205)	(0.305)
First Stage			
Δw_{t-1}	-0.021***	-0.020***	-0.021***
	(0.004)	(0.039)	(0.004)
$W * \Delta w_{t-1}$	-0.070***	-0.105***	-0.093***
	(0.007)	(0.006)	(0.012)
Bartik	-0.259***	-0.254***	-0.256***
	(0.014)	(0.014)	(0.014)

Notes: IV estimation results of (II) when the clustering data CMZ 2020, LMA 1980 and LMA 1990 are used to specify the neighbouring matrix W. The top part of the table reports the results of the second stage of the IV procedure, while the bottom part reports the first stage. The data range is 1992-2021. The standard errors were clustered at the county level.
"***", "**" and "*" denote statistical significance at the 1, 5 and 10 per cent levels.

The tables above indicate that the coefficient for the local unemployment rate U_t ranges from -0.183 to -0.192 with OLS estimation and from -0.678 to -0.768 with IV estimation, depending on the specified W matrix. In comparison to the results presented in the Tables 2.5 and 2.6, these coefficients are larger than the ones obtained using a contiguity matrix. In contrast, the coefficients for $W * U_t$ exhibit an opposing trend, showing smaller absolute values than in the contiguity matrix scenario. Specifically, they range from -0.097 to -0.126 for OLS estimation and from -0.295 to -0.614 for IV estimation. In the IV case, coefficients lose significance for LMA 1990, but the estimated coefficient remains consistent with other results. This loss in significance can be attributed to the reduced precision associated with IV estimators.

The decrease in the coefficient on $W * U_t$ observed when transitioning from a contiguity matrix to those employing CMZ or LMA data can be explained by a more pronounced relationship between the labour markets of counties that share a physical border. As previously noted, using CMZs or LMAs to define the W matrix is preferable to the contiguity matrix, as they more effectively capture dynamics such as commuting patterns and other pertinent factors related to local labour markets, while allowing for counties that do not share a physical border to be related. Ignoring those counties can be problematic, as these regions may also offer viable employment opportunities for workers and are a source of labour for local firms. Labour market tightness in those counties would also impact the local wage level, and not accounting for it would lead to biased estimations.

Alternative Bartik-style instruments: tradeable sectors only

The Bartik-style instruments used in the baseline IV analysis are constructed using employment shares across all sectors of the economy. However, one possible concern is that including non-tradeable sectors may introduce endogeneity, as growth in these sectors can reflect local economic conditions. In contrast, tradeable sectors are more likely to be driven by national demand shocks and are therefore less susceptible to local labour supply factors. To address this concern, I re-estimate equation (II) using an alternative set of Bartik-style instruments constructed only from tradeable sectors.

Among the sectors available in the QCEW employment data, the following are classified as tradeable and used in constructing the instruments: "Natural Resources," "Manufacturing," "Information," "Financial Activities," and "Professional and Business Services." The corresponding estimation results are reported in Table 2.11.

Table 2.11: IV estimation results: CMZ and LMA neighbouring matrices (Bartik-style instruments - tradeable only)

Neighbouring Matrix	**Contiguity**	**CMZ 2020**	**LMA 1980**	**LMA 1990**
		Second Stage		
Δw_{t-1}	-0.192***	-0.187***	-0.184***	-0.185***
	(0.016)	(0.017)	(0.016)	(0.017)
$W * \Delta w_{t-1}$	-0.058***	-0.060***	-0.119***	-0.093***
	(0.017)	(0.018)	(0.024)	(0.036)
U_t	-1.072***	-1.270***	-1.355***	-1.333***
	(0.156)	(0.157)	(0.169)	(0.199)
$W * U_t$	-1.602***	-1.145***	-1.106***	-1.103***
	(0.173)	(0.182)	(0.214)	(0.324)
		First Stage		
Δw_{t-1}	-0.022***	-0.023***	-0.022***	-0.023***
	(0.005)	(0.005)	(0.005)	(0.005)
$W * \Delta w_{t-1}$	-0.093***	-0.072***	-0.105***	-0.093***
	(0.006)	(0.008)	(0.007)	(0.012)
Bartik	-0.178***	-0.180***	-0.176***	-0.178***
	(0.011)	(0.011)	(0.011)	(0.011)

Notes: IV estimation results of (II) using the clustering data CMZ 2020, LMA 1980 and LMA 1990 to specify the neighbouring matrix W. In contrast to the IV results presented in Table 10, the Bartik-style instruments used for this estimation include only tradeable sectors. The top part of the table reports the results of the second stage of the IV procedure, while the bottom part reports the first stage. The data range is 1992-2021. The standard errors were clustered at the county level.
"***", "***" and "*" denote statistical significance at the 1, 5 and 10 per cent levels.

As can be observed in the table above, the estimates obtained with the tradeable-sector instruments imply a steeper wage Phillips curve than the baseline OLS results and are also larger in absolute value than the previous IV estimates using all sectors. The coefficients on local unemployment are around -1.3 when a CMZ neighbour matrix is used, while the estimated spillover coefficients are approximately -1.1in the same specifications. These results, therefore move further away from the OLS estimates and indicate a stronger sensitivity of wage inflation to both local and neighbouring labour-market slack.

This pattern is consistent with the concern that the inclusion of non-tradeable sectors in the shift-share instruments may introduce local endogeneity or attenuation, biasing the estimated WPC slope towards zero. When the instruments are restricted to tradeable sectors, the identifying variation is more plausibly driven by national sectoral shocks rather than local labour-market conditions. The results are therefore supportive of the main conclusion of the chapter. Once endogeneity and spatial spillovers are accounted for, the estimated wage Phillips curve is substantially steeper than suggested by non-instrumented specifications.

2.3.3 Aggregation to a nation-level curve

This section presents the following analysis step, which shows the aggregation of the regional WPC with spatial spillovers estimated above into a nation-level WPC. I depart from the preferred specification (II) for a generic county and aggregate it into a nation-level WPC. Take the i^{th} line from the system defined by (II):

$$\Delta w_{i,t} = \alpha_i + \delta_t + \beta_1 \Delta w_{i,t-1} + \beta_2 W_i \Delta \boldsymbol{w_{t-1}} + \gamma_1 U_{i,t} + \gamma_2 W_i \boldsymbol{U_t} + \mu_{i,t} \tag{2.7}$$

W_i is the i^{th} row from the spatial weights matrix W. As W is row-normalized, it follows that the sum of the elements of W_i are equal to 1 and $W_i \boldsymbol{U_t}$ is the average unemployment rate in the neighbourhood of county i.

Define $\tau_{i,t}$ as a regional share of the national labour force in year t, such that $\sum_i \tau_{i,t} = 1$. The nation level wage inflation (Δw_t) and unemployment rate (U_t) are denoted by dropping the regional index i. They are given by the averages weighted by $\tau_{i,t}$:

$$\Delta w_t = \sum_i \tau_{i,t} \Delta w_{i,t}, \tag{2.8}$$

$$U_t = \sum_i \tau_{i,t} U_{i,t}. \tag{2.9}$$

Multiplying (2.7) by $\tau_{i,t}$:

$$\tau_{i,t}\Delta w_{i,t} = \tau_{i,t}\alpha_i + \tau_{i,t}\delta_t + \tau_{i,t}\beta_1 \Delta w_{i,t-1} + \tau_{i,t}\beta_2 W_i \boldsymbol{\Delta w_{t-1}} + \tau_{i,t}\gamma_1 U_{i,t} + \tau_{i,t}\gamma_2 W_i \boldsymbol{U_t} + \tau_{i,t}\mu_{i,t}$$

Summing across all regions:

$$\sum_i \tau_{i,t}\Delta w_{i,t} = \sum_i \tau_{i,t}\alpha_i + \delta_t \sum_i \tau_{i,t} + \beta_1 \sum_i \tau_{i,t}\Delta w_{i,t-1} + \beta_2 \sum_i \tau_{i,t} W_i \boldsymbol{\Delta w_{t-1}}$$
$$+ \gamma_1 \sum_i \tau_{i,t} U_{i,t} + \gamma_2 \sum_i \tau_{i,t} W_i \boldsymbol{U_t} + \sum_i \tau_{i,t}\mu_{i,t}.$$

Denoting $\alpha = \sum_i \tau_{i,t}\alpha_i$, $\mu_t = \sum_i \tau_{i,t}\mu_{i,t}$, and $\boldsymbol{T_t}$ as a (1xN) vector stacking all τ at time t, and using (2.8) and (2.9), the equation above can be written as:

$$\Delta w_t = \alpha + \delta_t + \beta_1 \Delta w_{t-1} + \beta_2 \boldsymbol{T_t} W \boldsymbol{\Delta w_{t-1}} + \gamma_1 U_t + \gamma_2 \boldsymbol{T_t} W \boldsymbol{U_t} + \mu_t. \tag{2.10}$$

Notice that $\boldsymbol{T_t}W$ is a (1xN) vector that interacts the spatial weights with the labour force shares. As the rows of $\boldsymbol{T_t}$ and W sum to one, the elements of $\boldsymbol{T_t}W$ also sum up to one, so when it pre-multiplies $\boldsymbol{U_t}$, it results on a weighted average for the unemployment rate. This average takes labour force shares into consideration but also accounts for the centrality of each county: the more neighbours a region has, the higher its weight on the determination of the average. In addition, regions that border areas with few or no other neighbours receive a higher weight due to the row normalization of W.

Now, decompose region's i unemployment rate into two components, the average (national) unemployment rate (U_t), and the deviation from the average ($\tilde{U}_{i,t}$):

$$U_{i,t} = U_t + \tilde{U}_{i,t}.$$

The vector of regional unemployment $\boldsymbol{U_t}$ can then be decomposed in two parts, one that repeats the average unemployment rate U_t N times, and another that stacks $\tilde{U}_{i,t}$, which is defined as $\tilde{\mathbf{U}}_t$. Now denote $\tilde{U}_t^{TW} = \boldsymbol{T_t} W \tilde{\mathbf{U}}_t$. The variable $\tilde{U}_t^{TW}$ represents a weighted average measure of the regional unemployment rate deviation from the national average. Notice that $\tilde{U}_t^{TW}$ is not necessarily zero as it used weights from the matrix $\boldsymbol{T_t} W$ and not from $\boldsymbol{T_t}$. $\boldsymbol{T_t} W$ can be interpreted as a size-weighted centrality measure. It gives more importance to some region over others, as it takes into account how connected each region is along with the size of their labour force. The element $\boldsymbol{T_t} W \boldsymbol{U_t}$ in (2.10) can be re-written as:

$$\boldsymbol{T_t} W \boldsymbol{U_t} = U_t + \tilde{U}_t^{TW}.$$

Following a similar procedure for the lagged wage inflation, we have that:

$$\boldsymbol{T_t} W \Delta \boldsymbol{w_{t-1}} = \Delta w_{t-1} + \Delta \tilde{w}_{t-1}^{TW}.$$

Replacing the expressions above in (2.10) gives:

$$\Delta w_t = \alpha + \delta_t + (\beta_1 + \beta_2) \Delta w_{t-1} + \beta_2 \Delta \tilde{w}_{t-1}^{TW} + (\gamma_1 + \gamma_2) U_t + \gamma_2 \tilde{U}_t^{TW} + \mu_t. \tag{2.11}$$

Equation (2.11) presents the final outcome of the aggregation process. The unemployment rate appears twice in the equation: first as aggregate unemployment rate U_t, and second as the deviation from this aggregate ($\tilde{U}_t^{TW}$). In the event of an aggregate shock, only U_t will be affected. Since this shock influences all counties equally, it leads to no variation in $\tilde{U}_t^{TW}$. This means that the coefficient that measures the impact of the national level of unemployment on the wage inflation is $\gamma_1 + \gamma_2$.

Calculating the coefficient that controls for the effect of aggregate shocks from the estimation results gives coefficients that range from -0.292 to -0.325 for the OLS estimation and from -1.063 to -1.657 when IV with Bartik-style instruments procedure is used. The coefficient on the aggregate unemployment rate indicates that the total effect of an aggregate shock is the combination of the local impact of the shock and the average spillover effect on neighbouring regions, which is reflected in the sum of both coefficients.

In contrast, the component $\tilde{U}_t^{TW}$ captures the effects of idiosyncratic shocks on the national wage inflation. As noted earlier, this component measures deviations of the unemployment rate from the national average, representing a weighted average of these deviations. It is weighted by the matrix $\boldsymbol{T_t}W$, which assigns greater significance to countries with higher centrality. Centrality is determined by two factors: first, a region becomes more central as the number of neighbouring areas increases; second, a region's centrality is enhanced when its neighbours have fewer nearby regions, thereby reducing competition from adjacent counties. This feature results from the normalization of the rows of W.

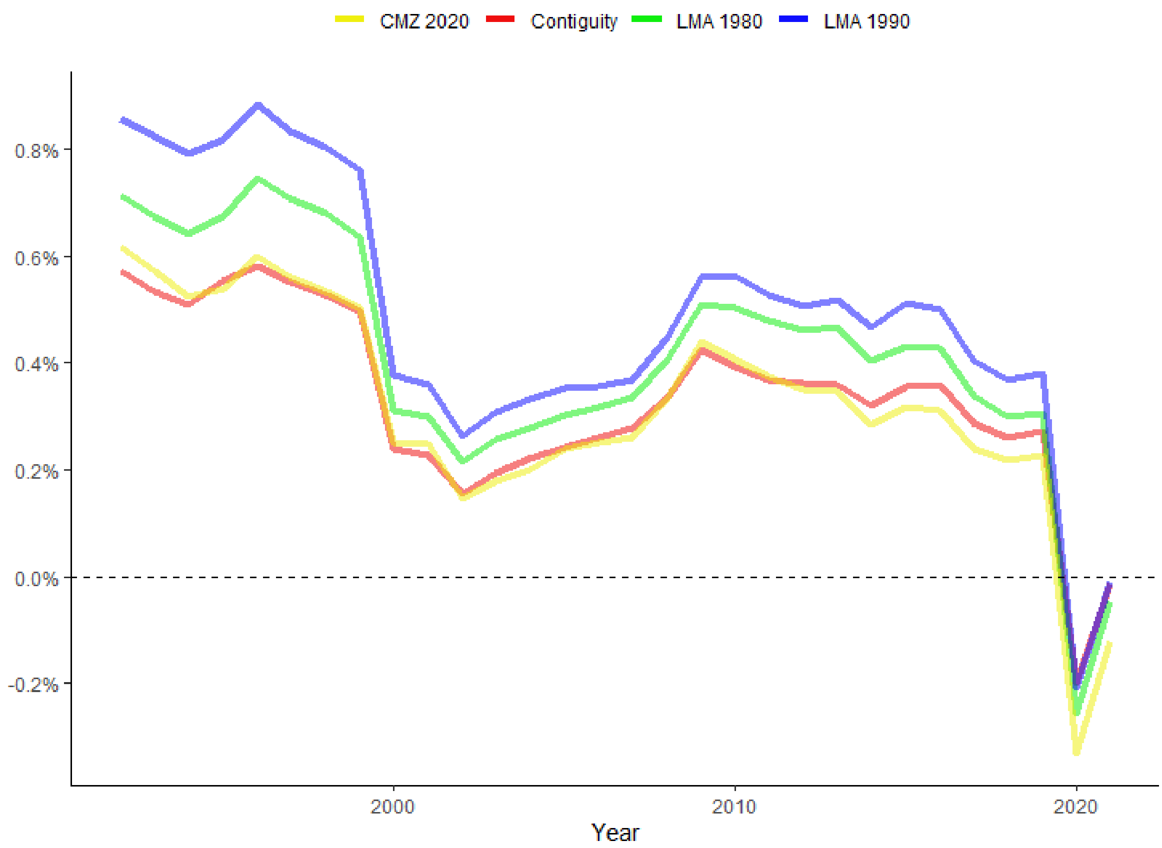
The result described above has significant policy implications. It suggests that better-connected counties should receive extra attention in policy design, as shocks impacting these areas will have a more pronounced effect on aggregate wage inflation than average due to stronger spillover effects. Since these regions experience commuting flows from various areas, an increase in local labour market tightness means that firms in those counties will be competing for workers from all the connected regions. This competition can lead to rising wage inflation across multiple locations.

Assuming that γ_2 is negative as in the empirical exercise, if the unemployment rate in the regions with higher centrality exceeds the national average ($\tilde{U}_t^{TW} > 0$), wage inflation should be lower than what is predicted by looking at the coefficient on the national level unemployment rate ($\gamma_1+\gamma_2$), requiring a correction by looking at the deviation component $\tilde{U}_t^{TW}$.

Figure 2.10 displays the implied $\tilde{U}_t^{TW}$ over time for the different neighbouring matrices used in the empirical exercise in the last section. It is noteworthy that the values remain above zero throughout the entire data range, with the exception of the years 2020 and

2021, which were likely impacted by the COVID-19 pandemic. This suggests that regions exhibiting higher centrality consistently have unemployment rates above the national average, which implies that until 2020, the effect of unemployment rates on wage inflation was higher than the one predicted by $U_t * (\gamma_1 + \gamma_2)$.

Figure 2.10: Unemployment rate: weighted average deviation from national average ($\tilde{U}_t^{TW}$)



Notes: *The figure displays the computed values of $\tilde{U}_t^{TW}$ from 1992 to 2021. Each line corresponds to a series computed using a different neighbouring matrix W, as noted in the legend.*

2.3.4 Sample splitting estimation

This subsection aims to contribute to the existing literature on the stability of the WPC over time, which includes works such as Galí and Gambetti (2019). I evaluate the stability of the WPC following the global financial crisis by employing a sample splitting estimation approach. The estimation follows the same procedure as in the previous section, but the sample is divided into two periods: the years leading up to the crisis (1992-2006) and the years encompassing and following the crisis (2007-2021). Each subsample comprises an equal number of years and, consequently, the same number of observations. Table 2.12 presents the OLS estimation results, while Table 2.13 displays the IV estimation results using shift-share instruments. For the second subsample (2007-2021), the Bartik instruments were recalibrated and use the sectoral employment shares for 2007, the initial year of this subsample.

Table 2.12: Estimation results: Sample splitting estimation (OLS)

Time Frame	**1992 - 2006**			**2007 - 2021**		
Neighbouring Matrix	**CMZ 2020**	**LMA 1980**	**LMA 1990**	**CMZ 2020**	**LMA 1980**	**LMA 1990**
Δw_{t-1}	-0.191***	-0.191***	-0.191***	-0.103***	-0.104***	-0.104***
	(0.015)	(0.015)	(0.015)	(0.016)	(0.012)	(0.016)
$W * \Delta w_{t-1}$	0.075***	0.097***	0.086***	0.119***	0.152***	0.160***
	(0.015)	(0.014)	(0.021)	(0.013)	(0.015)	(0.015)
U_t	-0.118***	-0.127***	-0.122***	-0.350***	-0.370***	-0.365***
	(0.033)	(0.032)	(0.033)	(0.035)	(0.035)	(0.035)
$W * U_t$	-0.171***	-0.153***	-0.172***	0.008	0.050	0.046
	(0.032)	(0.033)	(0.035)	(0.035)	(0.036)	(0.035)

Notes: OLS estimation results of (II) with sample splitting in two periods: pre-crisis (1992-2006) and after-crisis (2007-2021). The clustering data CMZ 2020, LMA 1980 and LMA 1990 are used to specify the neighbouring matrix W. The standard errors were clustered at the county level.
"***", "**" and "*" denote statistical significance at the 1, 5 and 10 per cent levels.

Table 2.13: Estimation results: Sample splitting estimation (Bartik-style instruments)

Time Frame	**1992 - 2006**			**2007 - 2021**		
Neighbouring Matrix	**CMZ 2020**	**LMA 1980**	**LMA 1990**	**CMZ 2020**	**LMA 1980**	**LMA 1990**
			Second Stage			
Δw_{t-1}	-0.210***	-0.209***	-0.209***	-0.125***	-0.123***	-0.121***
	(0.160)	(0.158)	(0.016)	(0.014)	(0.013)	(0.014)
$W * \Delta w_{t-1}$	0.006	0.000	0.009	0.036	0.024	0.050
	(0.017)	(0.019)	(0.026)	(0.023)	(0.034)	(0.034)
U_t	-1.125***	-1.182***	-1.176***	-0.600***	-0.604***	-0.631***
	(0.213)	(0.221)	(0.228)	(0.204)	(0.215)	(0.213)
$W * U_t$	-0.650***	-0.508**	-0.525*	-0.661***	-0.672***	-0.509**
	(0.211)	(0.252)	(0.300)	(0.184)	(0.232)	(0.213)
			First stage			
Δw_{t-1}	-0.010***	-0.009***	-0.012***	-0.027***	-0.025***	-0.025***
	(0.003)	(0.003)	(0.004)	(0.002)	(0.002)	(0.002)
$W * \Delta w_{t-1}$	-0.033***	-0.050***	-0.045***	-0.082***	-0.119***	-0.119***
	(0.006)	(0.006)	(0.011)	(0.004)	(0.005)	(0.005)
Bartik	-0.293***	-0.289***	-0.429***	-0.303***	-0.297***	-0.297***
	(0.030)	(0.030)	(0.048)	(0.016)	(0.016)	(0.016)

Notes: IV estimation results of (II) with sample splitting in two periods: pre-crisis (1992-2006) and after-crisis (2007-2021). The clustering data CMZ 2020, LMA 1980 and LMA 1990 are used to specify the neighbouring matrix W. The top part of the table reports the results of the second stage of the IV procedure, while the bottom part reports the first stage. The standard errors were clustered at the county level.
"***", "**" and "*" denote statistical significance at the 1, 5 and 10 per cent levels.

The findings presented in Table 2.12 indicate that the coefficient of the aggregate WPC remains relatively stable across both subsamples. However, they also suggest that in the post-crisis period, the spatial spillover effects do not account for the level of wage inflation. This is evidenced by the loss of statistical significance of the coefficient on $W * U_t$, while the magnitude of the coefficient on U_t itself increases.

The findings presented in Table 2.13 highlight significant differences. They indicate a flatter aggregate WPC in the post-crisis period, with the coefficient shifting from a range of -1.690 to -1.775 in the subsample from 1992 to 2006, to a range of -1.140 to -1.276 in the subsample from 2007 to 2021. Moreover, a considerable portion of the decline in the aggregate coefficient is attributed to a substantial reduction in the estimated coefficient on U_t. Notably, the direction of the IV results contrasts with that of the OLS, where the coefficient on U_t increases while that on $W * U_t$ decreases. This discrepancy can be explained by the differing severity of the endogeneity issue across the various time periods, resulting in distinct biases in OLS estimation for each subsample.

2.3.5 Commuting flow neighbouring matrices

This subsection extends the analysis by constructing a spatial weight matrix directly from inter-county commuting flows. This approach has two advantages over the use of CMZs and LMAS. First, it is more flexible, as it does not limit the county to be neighbour only to the other counties in the same block. Second, it does not assign equal weights to all neighbouring counties. Instead, weights are proportional to the observed intensity of commuting flows between county pairs.

The commuting data used to construct the new W matrix is sourced from American Community Survey (ACS) commuting-flow files, available every five years from 2006 to 2020. For earlier years, I use the corresponding commuting-flow data from the 1990 and 2000 U.S. Decennial Censuses[15]. As the estimation data sample ranges from 1990 to 2021, I construct a time-invariant matrix using the average commuting flow between county pairs over the 1990 to 2020 period.

As in the other specifications in this chapter, the resulting W matrix is row-normalised. However, instead of defining neighbours through binary indicators, the entries of the matrix are based on the share of total commuting flows associated with each county. The commuting flow between counties i and j is defined as a two-way flow, combining both workers who live in the county i and work in the county j, and workers who live in the county j and work in the county i. The corresponding OLS and IV results are reported in the table below:

[15] All data mentioned above is available at `https://www.census.gov/topics/employment/commuting/guidance/flows.html`

Table 2.14: Estimation results: Commuting flow-based neighbouring matrices

	OLS	**IV**
Δw_{t-1}	-0.142*** (0.015)	-0.168*** (0.017)
$W * \Delta w_{t-1}$	0.134*** (0.019)	0.093*** (0.022)
U_t	-0.144*** (0.021)	-1.012*** (0.156)
$W * U_t$	-0.216*** (0.026)	-1.987*** (0.235)
First stage (IV)		
Δw_{t-1}	-0.022*** (0.005)	
$W * \Delta w_{t-1}$	-0.099*** (0.010)	
$Bartik$	-0.178*** (0.011)	

Notes: The table presents the estimation results of (II) for when commuting flow data is used to specify the neighbouring matrix W. The top panel of the table provides the OLS and IV estimation results, while the bottom panel of the table reports the first stage for the IV approach. The data range is 1992-2021. The standard errors were clustered at the county level. "***", "**" and "*" denote statistical significance at the 1, 5 and 10 per cent levels.

Compared to the results in Tables 2.9 and 2.10, which use the other specifications of the spatial weight matrix, the commuting-flow matrix generates larger estimated spillover effects. In the OLS specification, the coefficient on neighbouring unemployment increases in absolute value relative to the previous estimates, reaching -0.216. In the IV specification, the corresponding spillover coefficient becomes substantially larger, reaching -1.987. These results suggest that commuting flows capture a more direct form of labour-market interaction than block-based definitions such as CMZs and LMAs.

However, these strong spillover results should be interpreted with caution, as they may partly reflect the large number of counties with commuting flows that are sizeable relative to their own local labour force. In the data used to construct the W matrix, many counties display substantial commuting flows relative to their within-county employment flows. When considering one-way commuting flows, approximately 23% of counties have outflows to other counties that exceed their within-county flow, while around 53% have outflows equal to at least 50% of their within-county flow. These are often small counties in terms of labour-market size. With these facts in mind, treating all counties equally in the estimation procedure may inflate the estimated spillover effects from neighbouring unemployment. Therefore, future research should consider specifications that explicitly account for county size.

2.4 Concluding remarks

In this chapter, I draw on county-level labour market data to estimate a wage Phillips curve. This approach exploits the advantages of using disaggregated data. It offers a significantly greater number of observations and increased variability. Additionally, it allows for a panel structure that can account for unobserved variables through fixed effects. The regionally disaggregated data used in this chapter are at the county level, covering the entirety of the country area, allowing for an analysis of the spatial dynamics within local labour markets, which is a key contribution of this study.

The results present evidence of spatial spillovers occurring in local labour markets within the framework of a regional wage Phillips curve. The labour market conditions of neighbouring areas help explain local wage inflation, so excluding them from the estimated specification of the regional WPC may result in biased estimates of the curve coefficients.

The spillover effects can be attributed to the considerable number of workers in the US who commute for work. Strong commuting patterns suggest that companies can seek and hire workers not only within their own county but also from surrounding areas. Consequently, the tightness of the local labour market may lead to increased wages in neighbouring regions, as local firms compete for workers from these locations.

This study also demonstrates that labour market areas (LMAs) are well-suited as neighbouring matrices when employing a spatial approach to estimate the wage Phillips curve. Since LMAs are derived from commuting data, they effectively capture the spillover effects generated within the labour market. This is expected, as commuting is anticipated to be the primary reason for such spillover effects. A key indicator of that result is that when the data is aggregated at the LMA level, the spillover effects from unemployment rates into wage inflation disappear, suggesting that these effects are contained within the LMAs.

I have also discussed how the regional wage Phillips curve with spatial spillover effects can be mapped into a national-level curve. The aggregation exercise reveals that the coefficient reflecting the influence of the national unemployment rate on national wage inflation is composed of two components: the coefficient that captures the local effect of unemployment rates on wage inflation, along with the coefficient that represents the spillover effect of unemployment rates from neighbouring areas. The aggregate curve also encompasses a component that assesses the influence of idiosyncratic shocks on the national wage inflation rate. It underscores that counties with larger labour forces and a greater number of neighbouring areas experience a more pronounced effect from local shocks on the national level of wage inflation.

The estimated coefficients for the national-level WPC, when an IV approach and shift-share instruments are used, range between -1.063 and -1.657, depending on the clustering data used to determine the neighbouring matrix. A considerable share of those coefficients are determined by the labour markets spillovers. Those results indicate a high sensitivity of wage growth to labour market tightness. This is crucial for modelling inflation-unemployment dynamics, as it shows that pressures in the labour market lead to higher production costs via elevated wages.

The larger coefficient magnitudes arise once spatial features are incorporated, which are typically ignored in standard estimation approaches. This underscores the importance of accounting for spatial spillovers in wage setting, especially when the analysis departs from a regional or sectoral framework to a macro level. For example, when firms in one county face stronger demand and expand hiring, they draw workers from surrounding areas, bidding up wages beyond the county's borders. Such channels should be incorporated when mapping regional evidence into macro models and for policy design. An analogous logic could also apply in the sectoral dimension: shocks in one industry can propagate through input–output linkages, affecting wages in upstream suppliers and related downstream sectors.

A further key policy relevant finding of this chapter is that regional idiosyncratic shocks can have disproportionate effects on national wage inflation. The aggregation exercise shows that a local shock's contribution depends on a size-adjusted centrality measure: counties with many and/or stronger links to neighbours transmit shocks more widely and thus exert greater influence on aggregate wage growth. Holding population constant, two counties hit by shocks of equal magnitude can generate different national effects if their network positions differ; the more connected county produces larger spillovers to surrounding labour markets and, in turn, a larger aggregate impact. The policy implication is clear: monitoring and evaluation should prioritise high-centrality nodes in the wage-setting network, as shocks in these regions are more likely to propagate and amplify through spatial spillovers.

In conclusion, estimating the spatially augmented regional WPC and subsequently aggregating it to the national level offers several advantages. It reveals the unbiased coefficients related to the local level of unemployment rates, highlighting the effects of spatial spillover effects, and illustrates how idiosyncratic shocks impact national wage inflation, weighted by the significance of each county in terms of size and its connectivity to other labour markets. Together, these features provide a coherent micro-to-macro bridge for analysis and policy design.

Chapter 3

A procurement-based dataset on the US military expenditure

3.1 Introduction

Government expenditure increases often translate into boosted economic activity, leading to higher aggregated demand, impacting key variables such as GDP, unemployment, and inflation. Changes in expenditure may also lead to sectoral and geographical effects, triggering the reallocation of labour, capital, and other resources. Identifying the effects of government expenditure on economic activity is challenging due to the endogenous nature of the problem. While increased spending stimulates output, higher output generates more tax revenue, enabling further expenditure.

Using military expenditure data has become a popular approach in empirical research to address the simultaneity problem, replacing other commonly employed methods such as the shock identification procedure introduced by Blanchard and Perotti (2002). One advantage of using defence expenditure over general government expenditure is that it tends to be more exogenous. This is because military spending is typically influenced more by geopolitical factors, such as wars and conflicts, than economic conditions.

Nearly half of the American discretionary outlays are allocated to defence, and this expenditure is an important driver of American economic performance. As stated by Oatley (2015), American postwar military buildups are strongly correlated with economic booms. The author asserts that, as the country can easily attract foreign capital, increases in defence expenditure do not crowd out private investment. The opposite happens: debt-financed military expenditure generates a crowding in effect, boosting the economy even further.

An increase in military expenditure can also result in other effects besides boosting overall economic activity, such as structural changes in the economy. Increased relative spending in a sector may culminate in a cross-sectoral movement of workers and factors of production. A similar effect can happen at the geographical level if, for example, workers commute to areas that offer higher wages due to boosted labour demand arising from increased expenditure.

One of the seminal and most influential papers on using military expenditure data is

Ramey and Shapiro (1998). The authors introduce a method to isolate military expenditure shocks known in the literature as the "narrative approach." The procedure consists of reading narratives and historical accounts regarding military build-ups related to war events and identifying sharp increases in expenditures, interpreted as arguably exogenous expenditure shocks. Later on, other papers also used military data but followed different approaches. Hall (2009) uses national defence data sourced from national accounts, where it is categorised as a subcomponent of government expenditure. Fisher and Peters (2010) uses the stock returns of military contractors to identify government expenditure shocks.

While the use of military expenditure data in empirical macroeconomics is not new, in recent years the use of Department of Defence (DoD) procurement-based data has become increasingly popular in the literature, as its availability and quality have significantly improved over the past years. The main advantage of this data is the rich level of granularity it contains. It permits focusing on specific sectors, locations, or companies, which can be exploited to examine regional and sectoral impacts of government defence spending.

The use of procurement-based military expenditure data dates back to studies such as Hooker and Knetter (1994) and Hooker and Knetter (1997). In these studies, the authors investigate the impact of military expenditure on unemployment and economic activity, positioning themselves among the first to exploit the detailed granularity of procurement-based data. They emphasise that this type of data offers a level of variation that is often missing in aggregated datasets, pointing out that other forms of military expenditure data typically lack disaggregation down to the state or smaller administrative levels.

Over the last decade, the use of military procurement data has grown, particularly in studies that explore regional and sectoral variations. One influential work in the fiscal policy literature is Nakamura and Steinsson (2014), which leverages state-level variation in procurement-based military expenditure to estimate a fiscal multiplier. They calculate an 'open economy relative multiplier' for the American economy within the context of a fiscal and monetary union. This multiplier shows how GDP rises in states where relative spending has increased, isolating the effects from monetary and tax policy responses to a spending shock.

Dupor and Guerrero (2017) use procurement data to construct state-level measures of military expenditure and estimate income and employment multipliers. Demyanyk et al. (2019) use DoD procurement spending during the economic crisis (2007–2009) to investigate the relationship between consumer debt and fiscal multipliers. Their study examines how economic growth across different core-based statistical areas (CBSA) responds to fiscal stimulus under varying levels of consumer debt. Auerbach et al. (2020) further highlight the potential of procurement expenditure datasets, using their detailed information to estimate local fiscal multipliers. They construct expenditure measures at the city-sector level, enabling analysis of industry and geographical spillovers. Auerbach et al. (2022) extend this approach by using military procurement data to assess fiscal multipliers in relation to the severity of stay-at-home orders imposed on U.S. cities during the COVID-19 pandemic. In a recent study, Cox et al. (2024) analyse US federal procurement expenditure in detail, highlighting its granularity and heterogeneity.

Although the procurement data is not new, more detailed datasets have only become available in recent decades. In this chapter, I exploit the U.S. Department of Defence (DoD) procurement data, available through the `USAspending.gov` platform. One of the contributions of this chapter is the provision of new tools for conducting systematic and reproducible analyses of this rich dataset, demonstrated through a set of R codes developed to manage raw procurement data. Additionally, this chapter offers a descriptive overview of military procurement spending, aggregated across multiple dimensions, which may prove valuable for researchers investigating various policy or economic questions using this data source.

The remainder of this chapter is organised into three sections. Section 2 discusses the data source and outlines the key components of the raw procurement dataset, as well as the methodology and codes used to aggregate and analyse the data across various relevant dimensions. Section 3 presents a descriptive analysis of both the raw and aggregated procurement expenditure data. It begins by contextualising the data with pertinent geopolitical events, U.S. electoral cycles, and other factors that influence procurement and expenditure patterns. It then examines the distribution of the data across different dimensions, such as procurement value and length, while also analysing the comparison between obligations and actual spending and presenting a discussion of the evolution of the military expenditure deflator in relation to other price indices. Additionally, it analyses the distribution of expenditure in terms of sectoral, geographical, and firm-level dimensions. Section 4 concludes the chapter.

3.2 The procurement dataset

As mentioned in the introduction, this chapter presents a dataset on defence expenditure, constructed by aggregating procurement records from the `usaspending.gov` database. It also provides details on the aggregation procedure and the accompanying R code used in the process. This data source includes information on all procurement contracts awarded by the U.S. federal government since 2001. The data presented in this study are based on the contracts filtered by the agency, being restricted to the ones signed by the Department of Defence. The expenditure data consists of aggregated contract values across various dimensions of interest, such as time, location, and sector.

The procurement dataset is extensive, containing approximately 42 million contracts for the Department of Defence alone between 2001 and 2021. The data is highly detailed, including information such as the contract's signed date and the classification of the goods or services involved. The data also provides information about the company's location, distinguishing between the recipient's address and the place of performance. The latter is particularly important for capturing an increase in local demand, as it indicates where most funds are allocated in terms of wages and expenditures on intermediate goods and services.

3.2.1 Key components of the procurement records

The raw procurement database includes a vast list of components that may serve several purposes. For example, these components can be used to classify the types of products and services associated with the contract. Additionally, the database provides information about the companies involved, their locations, and various other factors. It also facilitates the identification of related contracts, as some may involve corrections or changes to previously signed procurement agreements. Below is a description of the key components of the raw database used in the compilation of the expenditure data:

Federal Action Obligation: This refers to the dollar amount that is either obligated or de-obligated through a procurement transaction. Essentially, it represents a commitment from the federal government to pay a specific amount, either immediately or in the future. If the obligation value is negative, it indicates a de-obligation, which reflects a modification or review of a previously signed contract. When the federal funds are paid out, this is referred to as an "outlay."

Procurement Instrument Identifier (PIID): A unique identifier assigned to a purchase contract, used for tracking the contract and any related modifications or transactions.

Parent Award Identifier (PAID): It consists of the PIID of the initial contract, being relevant to identify modifications to existing contracts. Not all PIIDs have PAIDs. If a given PIID has been reused (by multiple agencies), both the PIID and the PAID are required to identify a contract.

Data Universal Numbering System (DUNS) code: A nine-digit long number used to identify a business.

Unique Entity Identifier (UEI): The official identifier for companies that do business with the U.S. Government [1].

[1]From April 4, 2022, the federal government transitioned from using DUNS to UEI as the primary means of entity identification for federal awards government-wide.

North American Industry Classification System (NAICS): It is the standard system utilized by federal statistical agencies to classify business establishments. This classification aids in the collection, analysis, and publication of statistical data related to the U.S. business economy. The system employs a 6-digit code, where each digit represents a more detailed level of categorization [2].

Product Services Codes (PSC): This code categorises the products and services, including research and development (R&D), purchased by the federal government. The coding consists of 9 nine product codes, 23 service codes, and one R&D code. Product categories are represented by numbers from 1 to 9, while service codes are represented by letters. R&D services are indicated by the letter "A"[3].

Location components: City, County, State and ZIP code of the place of performance/recipient location[4].

Action date: The date the procurement was signed by the government.

Period of performance (start and end dates): The date from which the award is effective and the date it ends, respectively. The procurement length can be calculated as the time difference between these two dates. For the purposes of this chapter, I have calculated the procurement length on a monthly basis, determining the difference between the months of the start and end dates. If the dates fall within the same month, the contract is considered to have a length of one month.

3.2.2 Aggregating the data

The expenditure data presented in the next section were constructed by aggregating the reported "Federal Action Obligations" values from the DoD procurement contracts. The values were deflated using the implicit price deflator for government expenditure in

[2]For a list of the codes, please refer to `https://www.naics.com/six-digit-naics/`.

[3]The codes are defined in the Product and Service Codes Manual`https://www.acquisition.gov/psc-manual`.

[4]The county-level FIPS codes were not originally in the procurement dataset, but were merged into it using the location's county and state names.

national defence, sourced from the U.S. Bureau of Economic Analysis [5]. This deflator was selected as it more accurately represents the actual cost environment relevant to defence spending. The dataset includes the exact date each procurement contract was signed. However, since most of the values shown in this chapter are at the monthly frequency, each procurement value was adjusted for inflation using the corresponding monthly deflator.

One important aspect of the data is the length of the procurements. When taken into consideration, the results of data aggregation may vary significantly. For instance, if a $12,000 contract with a 12-month duration is signed in January 2020, disregarding the contract length would result in the entire $12,000 being attributed to January 2020. In contrast, if we consider the duration, the contract value would be spread out over time, assigning $1,000 for each month from January 2020.

Ignoring the contract length can lead to considerably different aggregated values, particularly in cases involving large and lengthy contracts. Overlooking the duration of contracts might be advantageous if the aim is to capture the anticipated impact of the expenditure at the time of signing. As Ramey (2011b) emphasises, the timing of economic shocks is crucial, as expectations regarding changes in expenditure may lead to different effects compared to unanticipated changes, which can have significant implications for econometric models. The aggregated procurement expenditure values reported in this chapter do not take the length into account in the data aggregation. However, subsection 3.3.7 presents the aggregate expenditure when procurement length is considered and compares and discusses the two approaches.

All data compilation and aggregation were performed using the software R. To achieve this, a collection of functions and code scripts was written to aggregate the raw data across different dimensions and filter it according to specified parameters, and also perform additional analysis. Appendix C reports the descriptions of all the R functions designed to aggregate the procurement expenditure. The scripts with the functions and additional codes are available at `https://github.com/Etexcs/Procurement-data-codes.git`[6]. Notice

[5]The deflator is available at the quarterly frequency, but it was converted to monthly using a linear interpolation.

[6]All the aggregation functions in the codebase, by default, do not account for contract duration. However, the script named “dissolved_data.R” performs an alternative aggregation procedure that dis-

the data files are not available at this repository as they exceed 100GB in size, but they can easily be downloaded from USAspending by filtering the procurements for the department of defence only.

The functions perform the aggregation of expenditure across various dimensions, including time (monthly or yearly), geography (national, state, county, or city), sector (based on NAICS or PSC codes), and company (identified by the UEI). In addition to aggregation, some functions also perform supplementary analyses, such as counting the number of procurements above a specified expenditure threshold, generating rankings of cities by expenditure within a given state, identifying the geographical distribution of expenditure within a specific county, or providing further disaggregation by NAICS categories.

3.3 Descriptive analysis of the data

This section presents a descriptive analysis of the DoD procurement data outlined in the previous section. I begin the section by contextualizing the procurement expenditure data within the framework of the American economy. Next, I examine the distribution of awarded procurements in terms of both value and duration. Following that, I show the distribution of the aggregated expenditure across various dimensions, including sectoral, geographical, and firm-level perspectives. I then discuss the behaviour of the aggregated expenditure when the contract lengths are considered in the aggregation process. I finish the section by comparing the price index used to adjust the military procurement expenditure data with other relevant price indexes.

tributes expenditure over the full length of each contract. This method is significantly more computationally intensive and was therefore parallelised to reduce runtime. Specifically, prior to aggregation, each procurement entry in the dataset is replicated for each month of its duration, with the contract value evenly split across those months. As a result, a single observation may generate multiple rows corresponding to the months over which the contract is active.

3.3.1 Military procurement expenditure in the context of the American economy

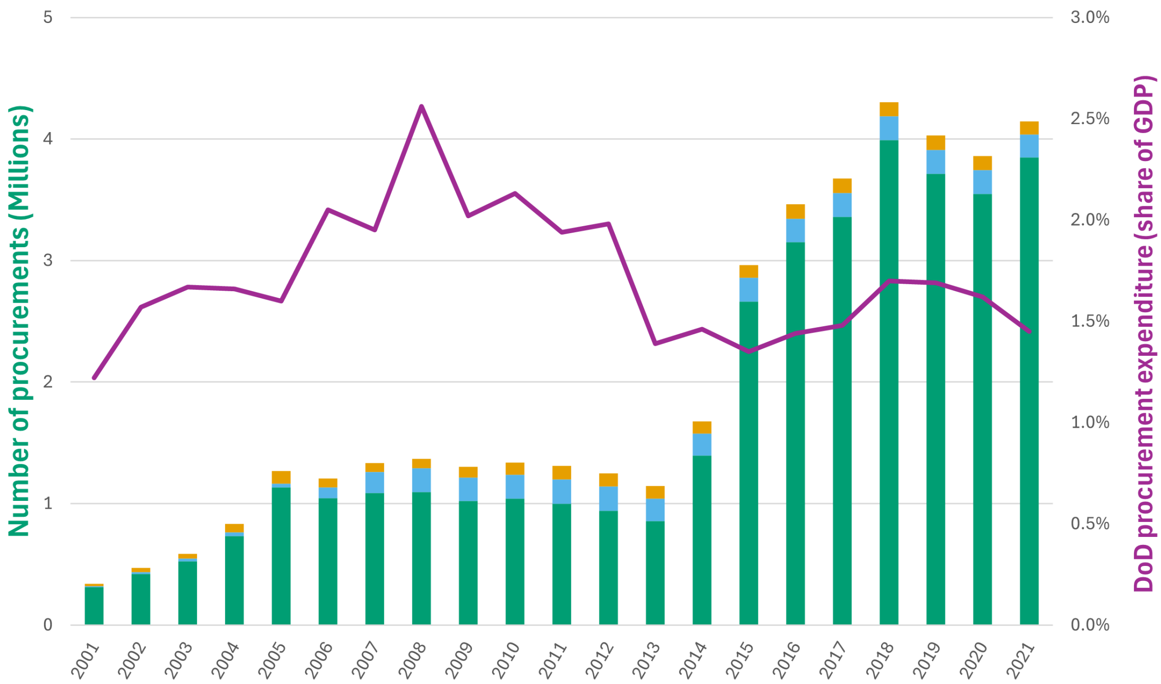
American government expenditure is classified as either mandatory or discretionary. Mandatory spending includes all forms of expenditure that are required by existing law, whereas discretionary spending requires annual budget approval from Congress and the president. Procurement expenditure is a form of discretionary spending. According to Cox et al. (2024), procurement expenditure accounts for 40% of federal government spending and 16% of total government spending, which is approximately 3% of the GDP. A significant portion of procurement expenditure is directed toward the Department of Defence, often exceeding half of the total awarded amount.

Therefore, the data referred to in the chapter as military, defence, or DoD procurement expenditure do not encompass the total military expenditure in the country; rather, they only reflect discretionary expenditure originating from contract obligations. Other defence expenditure data, such as pay and benefits for active and retired military personnel, fall under mandatory spending and are not included in the procurement expenditure measures discussed in this chapter.

Figure 3.1 displays the total number of contracts awarded each year in the raw DoD procurement database. It also shows the evolution of aggregated DoD procurement expenditure as a percentage of GDP[7]. The number of awarded contracts is represented in a bar plot: the green section indicates contracts with positive values, the light blue section represents contracts with zero values, and the orange section depicts de-obligations (negative values). All the observations in the dataset are represented in the bar plot, including new procurements modifications to existing ones.

[7]The GDP and government expenditure data (total) used in this section are sourced from the Bureau of Economic Analysis (BEA).

Figure 3.1: Awarded DoD procurements and total expenditure as a share of the GDP

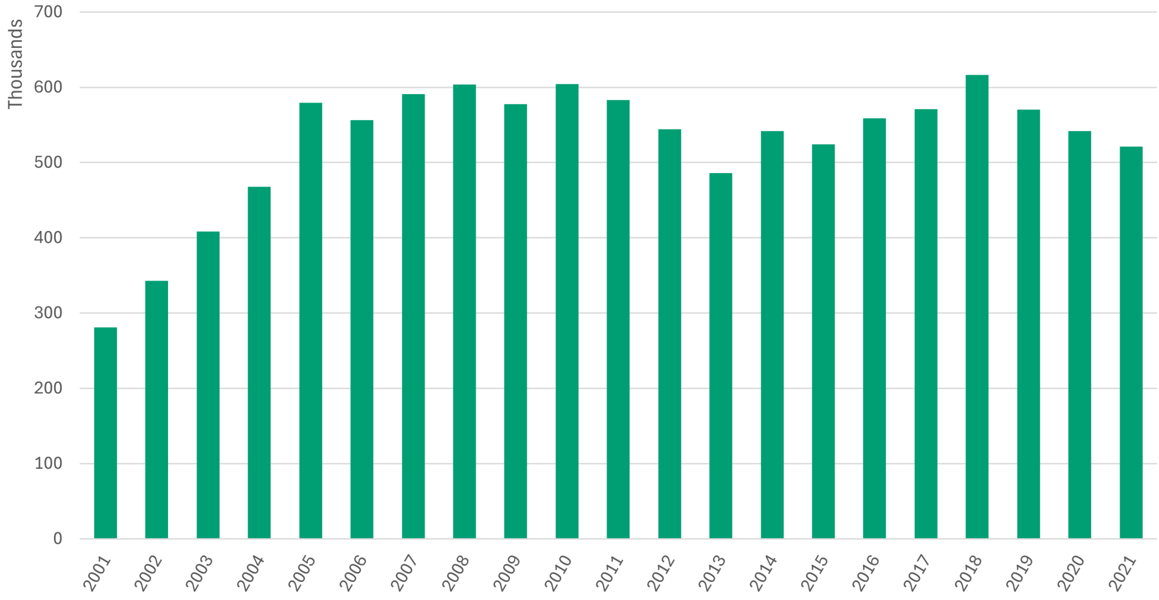


Notes: *The figure illustrates the total number of procurements awarded from 2001 to 2021, represented by a bar plot on the left axis. The green section of each bar indicates procurements awarded with positive values, while the light blue section represents procurements with a zero value, and the orange segment indicates procurements with negative values. Additionally, the figure shows the evolution of DoD procurement expenditure as a percentage of GDP, represented by the purple line on the right axis.*

The figure reveals that DoD procurement expenditure averages 1.7% of GDP within the analysed data range, with this percentage exceeding 2% from 2006 to 2010. There is a notable sharp increase in the number of observations starting in 2014; however, this rise is not matched by a corresponding increase in expenditure. This feature can be attributed to a change in data reporting standards, which has led to a decrease in the average value of reported procurements.

When focusing solely on procurements valued above $10,000, the contract numbers remain much more stable across the years, with their quantity better reflecting the increases in expenditure. This feature is illustrated in the figure below:

Figure 3.2: Yearly number awarded of DoD procurements with value over $10,000



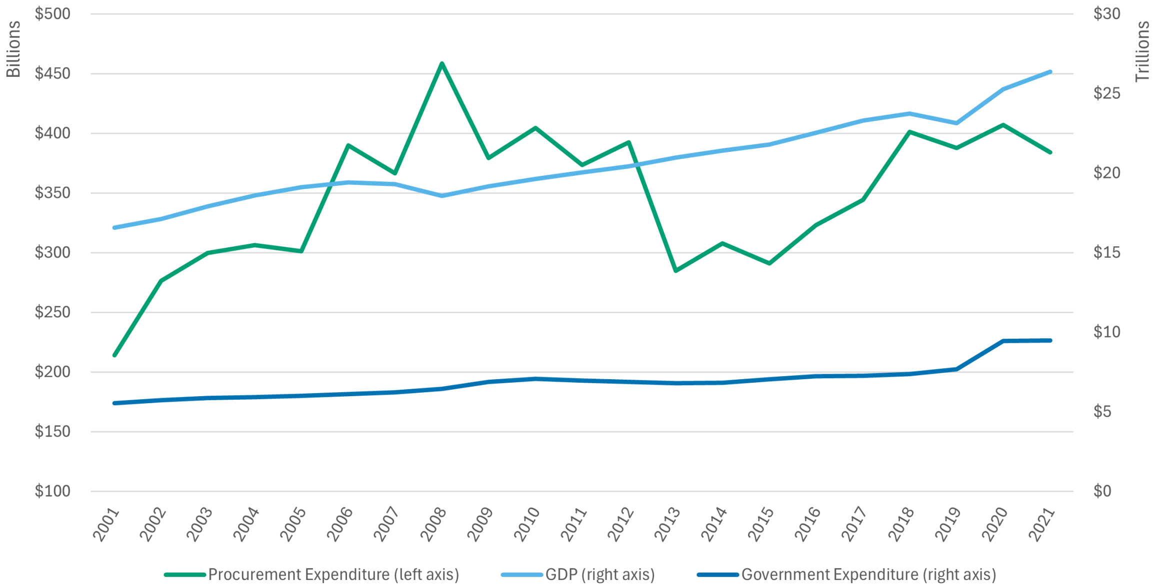
Notes: *The figure shows the total number of procurements awarded from 2001 to 2021, with a value exceeding ten thousand dollars.*

As illustrated in Figure 3.2, when contracts valued at less than $10,000 are excluded, the number of awarded procurements has remained stable since 2005. This is in contrast to Figure 3.1, where there is a significant increase in the number of procurements noted after 2015.

Figure 3.3 compares the behaviour of the DoD procurement expenditure with GDP and government expenditure from 2001 to 2021. Notably, the DoD procurement expenditure exhibits significantly greater volatility than the other two measures and does not indicate a clear relationship with GDP, aside from a general upward trend. For instance, the highest level of procurement expenditure was recorded in 2008, a year of recession. In contrast, total government expenditure demonstrates much more stability than military procurement expenditure.

As illustrated by Cox et al. (2024), federal government procurement purchases compose 16% of the total government expenditure, but represent around 48% of its total volatility. The authors also show that overall procurement expenditure is 3.8 times more volatile than the GDP. I proceed with a similar approach to that of Cox et al. (2024) to compare the volatility of the DoD procurement expenditure to the volatility of the GDP and the total government expenditure. Both the authors and this analysis use quarterly data. I first applied a Hodrick-Prescott (HP) filter to each data series[8]. Next, I calculated the standard deviation of the cyclical component and divided it by the mean of the trend component to establish a measure of volatility. This analysis indicates that DoD procurement expenditure is approximately 4.7 times more volatile than GDP and 2.2 times more volatile than total government expenditure. Notice that it surpasses the relative volatility found by the authors with respect to total procurement expenditure.

Figure 3.3: GDP, government expenditure and DoD procurement expenditure



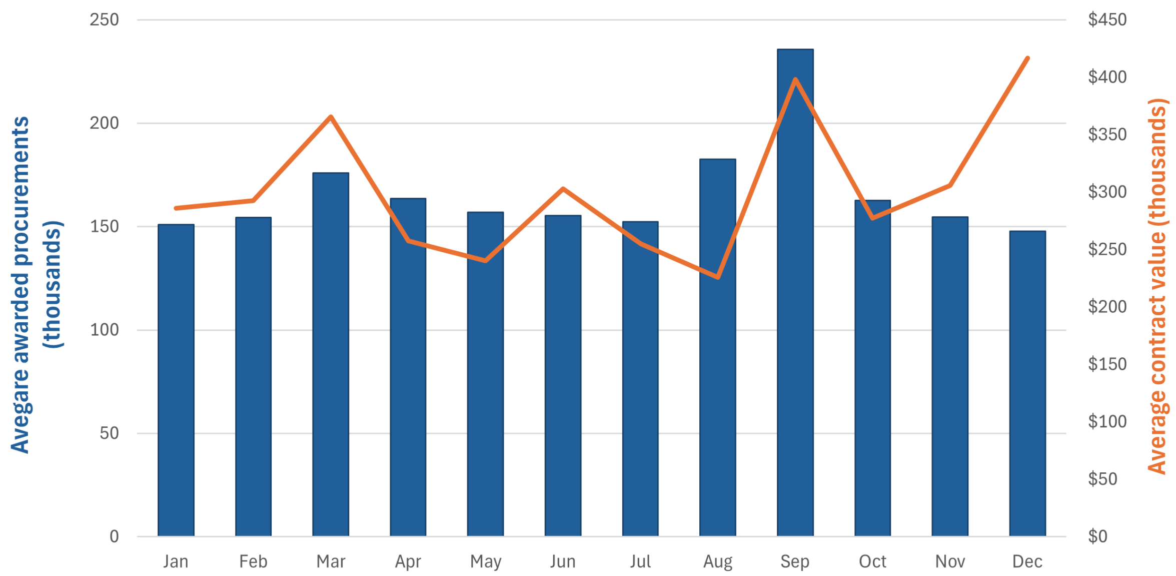
Notes: *The figure shows the changes in DoD procurement expenditure, GDP, and government expenditure from 2001 to 2021. The DoD procurement expenditure is represented on the left axis, while GDP and government expenditure are represented on the right axis of the figure. The DoD procurement was adjusted using the national defence government expenditure implicit price deflator, GDP was adjusted using the GDP implicit price deflator, and government expenditure was adjusted using the government expenditure implicit price deflator. All values are expressed in December 2021 dollars.*

As suggested by the figure above, the American DoD procurement expenditure can be associated to security-related events. Following the terrorist attacks of 2001, there was a significant increase in military spending, which continued until 2008 amid the intensification of the wars in Iraq and Afghanistan. In 2011, the war in Iraq came to an end, followed by a troop withdrawal from Afghanistan in 2012, the same period in which

[8] A smoothing parameter of 1,600 was applied.

a decrease in expenditure was observed. In 2013, a budgetary sequestration occurred due to the Budget Control Act of 2011, leading to further cuts in defence spending. These events, along with a reduction in U.S. operations in the Middle East, explain the sharp decline in defence procurement expenditure from 2008 to 2013. Starting in 2016, the resurgence of ISIS heightened perceptions of external threats. Additionally, there was a growing need to keep pace with China's expanding military capabilities, which resulted in an increase in defence expenditure from that year.

Figure 3.4: Monthly average number of awarded procurements and expenditure



Notes: *The figure illustrates the average number of procurements awarded each month, shown as a bar plot on the left axis. It also presents the average monthly procurement value represented by the orange line on the right axis. The data covers the years from 2001 to 2021. All procurement values have been adjusted using the national defence government expenditure implicit price deflator and are expressed in December 2021 dollars.*

Figure 3.4 illustrates the monthly average number of awarded contracts and their average values from 2001 to 2021. The data indicates that September consistently records a high average number of contracts. Furthermore, both September and December show elevated average contract values. These months represent significant spending periods, as they mark the end of the fiscal year and the calendar year, respectively.

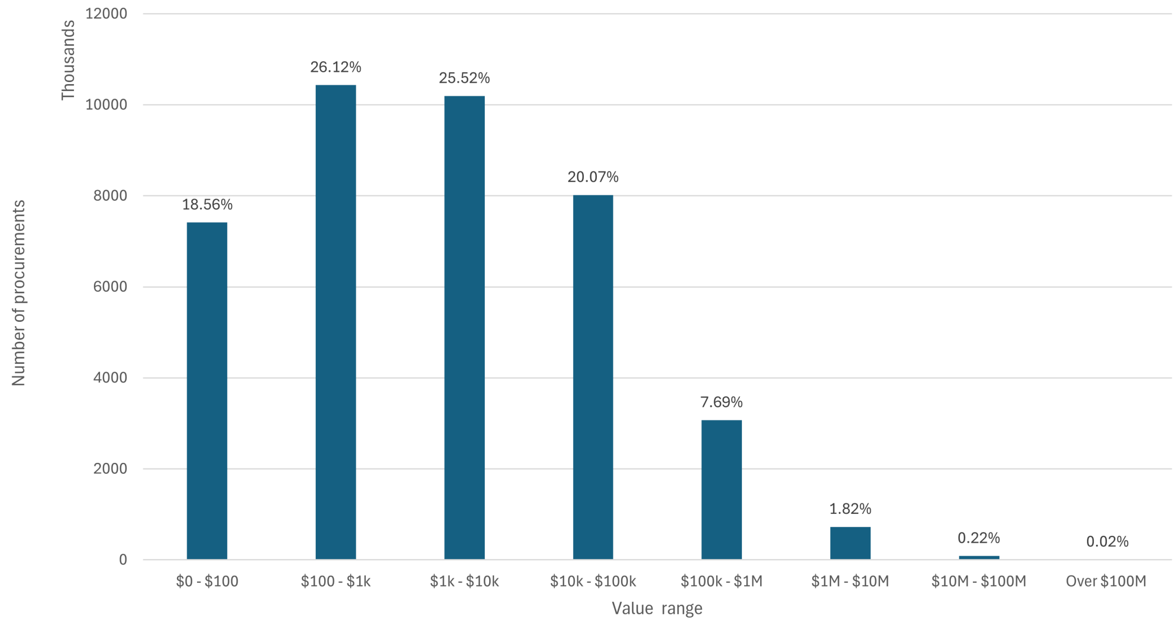
In September, agencies aim to utilize their remaining budget before it expires at the end of the fiscal year, as failing to do so may result in a reduced budget for the following year. Mintz (2002) notes that there is also a political aspect to the contracts signed in September, as they may be used as a tool to stimulate economic activity ahead of elections. Additionally, many high-value contracts are finalized in December, as numerous significant procurements are due for renewal at the end of the calendar year. Companies often prefer to establish year-long contracts in December to prepare for the upcoming year. Those numbers indicate then presence of a seasonality in the procurement expenditure numbers.

3.3.2 Distribution of awarded procurement values

This subsection analyses the distribution of procurements based on their value corrected for inflation, including all records in the database regardless of the year of signing. Figure 3.5 displays the number and percentage of contracts that fall within different value ranges. The figure does not include procurements with non-positive values. There is a significant volume of low-value contracts. According to the figure, over 70% of the contracts have values of $10,000 or lower[9]. In contrast, only about 2% of the data represents contracts with very high values exceeding one million dollars. Furthermore, there are 6,702 observations (0.02%) for contracts with values over 100 million dollars. The mean contract value is $199,665.20, while the median value is much smaller, at $1,753.64, indicating a very right-skewed distribution.

[9] As previously noted, a significant portion of the contracts with a value below $10,000 were registered after 2015, while in the years preceding it, only high-value procurements were registered in the database.

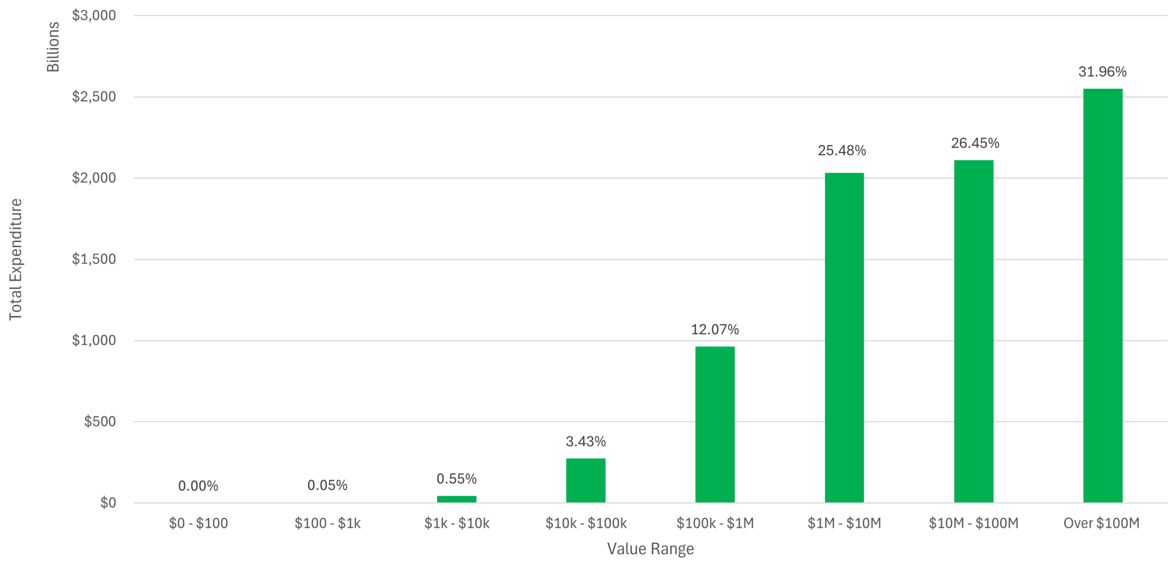
Figure 3.5: Distribution of DoD procurements per value range



Notes: *The figure displays the number of awarded procurements categorised by different value ranges from 2001 to 2021. All procurement values have been adjusted using the national defence government expenditure implicit price deflator and are expressed in December 2021 dollars. Contracts with zero or negative values have been excluded from the figure. The percentages displayed at the top of each bar represent the proportion of total awarded procurements that fall within each category.*

Figure 3.6 complements the analysis, showing how procurements in each category contributed to the total expenditure between 2001 and 2021. Although contracts valued over 100 million dollars account for only 0.02% of the total number of observations, they are responsible for nearly 32% of the total DoD expenditure. Additionally, contracts exceeding a million dollars represent almost 84% of the total expenditure. Therefore, figures 3.5 and 3.6 indicate that even though most procurements consist of small-value contracts, the majority of DoD procurement expenditure originates from a few high-value contracts.

Figure 3.6: Total DoD procurement expenditure per value range

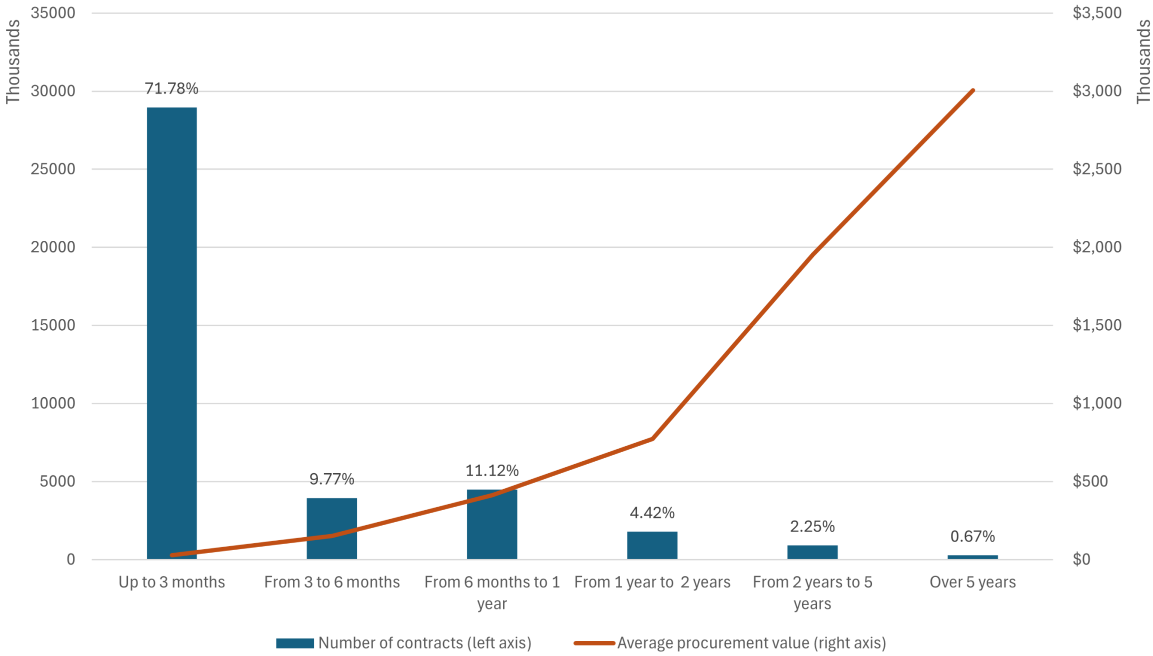


Notes: *The figure illustrates the total DoD procurement expenditure from contracts in different value ranges between 2001 and 2021. All procurement values have been adjusted for inflation using the national defence government expenditure implicit price deflator and are presented in December 2021 dollars. Contracts with negative values have been excluded from the figure. The percentages shown at the top of each bar indicate the proportion of total DoD procurement expenditure attributed to contracts within each category.*

3.3.3 Distribution by procurement length

This subsection provides a similar analysis by illustrating the distribution of the procurements based on their length. Figure 3.7 displays the number of awarded procurements categorized into length groups, along with the average contract value for each category. As the figure suggests, the majority of the raw database is composed of short and low-value contracts, with nearly 72% of the total procurements being shorter than 3 months and having an average value of $27,766.67.

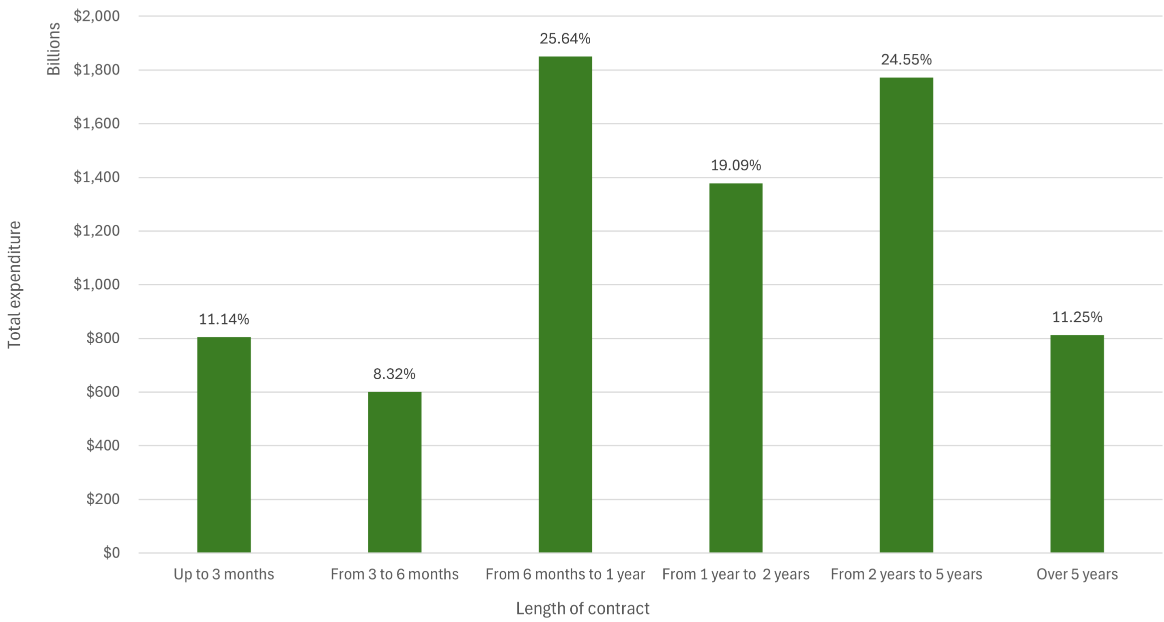
Figure 3.7: Number of procurements and average expenditure value per contract length group



Notes: *The figure illustrates the number of awarded DoD procurements from 2001 to 2021, categorised by different contract lengths. This information is represented through a bar plot on the left axis. The percentages displayed at the top of each bar indicate the proportion of the total awarded procurements that fall within each category. Additionally, the figure includes the average procurement value for each category, which is represented by the red curve on the right axis. All procurement values have been adjusted for inflation using the national defence government expenditure implicit price deflator and are presented in December 2021 dollars.*

Figure 3.8 displays the volume of expenditure derived from each length category. Although short-term procurements comprise nearly three-quarters of the total observations, they account for just over 11.1% of the total expenditure. Notably, over 90% of the observations fall within contracts of one year or less. However, around 55% of the total expenditure originates from contracts lasting more than one year, despite these representing only about 7% of all observations.

Figure 3.8: Total DoD procurement expenditure value per contract length group



Notes: *The figure illustrates the total DoD procurement expenditure from contracts in different length ranges between 2001 and 2021. All procurement values have been adjusted for inflation using the national defence government expenditure implicit price deflator and are presented in December 2021 dollars. The percentages shown at the top of each bar indicate the proportion of total DoD procurement expenditure attributed to contracts within each length category.*

Figures 3.6 and 3.8 indicate that a significant portion of the DoD procurement expenditure is attributed to long-term, high-value contracts. This suggests a considerable persistence in the military procurement expenditure. A similar finding is highlighted by Cox et al. (2024) but it encompasses all sources of procurements.

3.3.4 Sectoral distribution

This subsection discusses the sectoral distribution of the DoD procurement expenditure data. I first categorised the expenditure into different sectors using the North American Industry Classification System (NAICS) codes. Table 3.1 shows the shares of military procurement expenditures for the years 2005, 2010, 2015, and 2020, categorised by the 2-digit NAICS codes, which contain 21 categories and represent the broadest aggregation level of sectors within this classification system.

Table 3.1: Sectoral distribution of DoD procurement expenditure by NAICS 2-digit categories

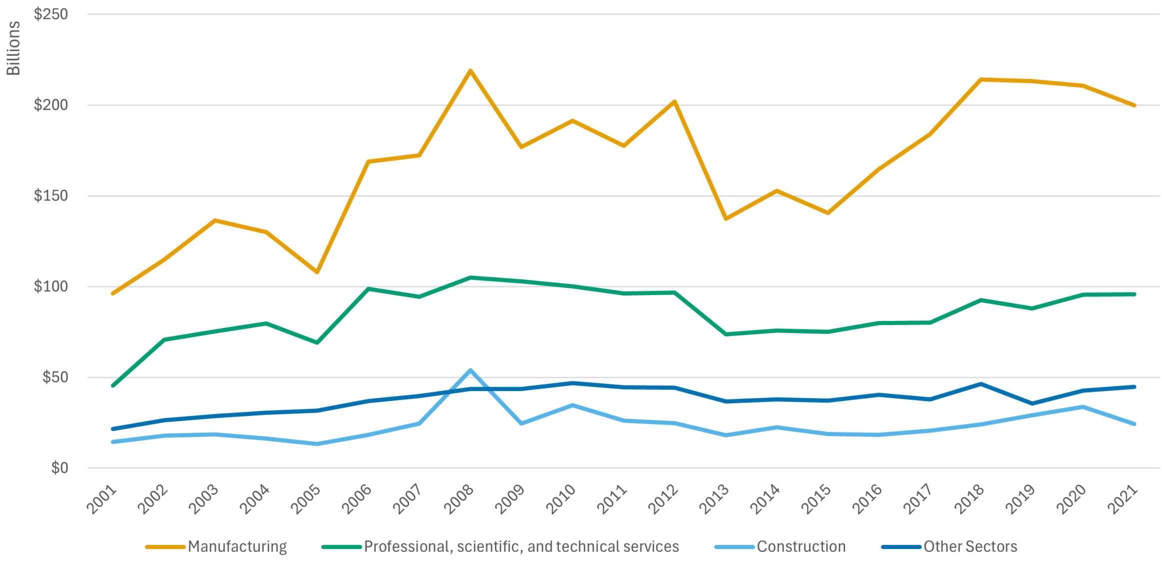
Sector by NAICS 2-digit	**Share of total expenditure**			
	2005	*2010*	*2015*	*2020*
Manufacturing	43.62%	47.41%	48.36%	51.75%
Professional, scientific, and technical services	27.87%	24.84%	25.80%	23.50%
Construction	5.33%	8.61%	6.50%	8.27%
Administrative, support, waste management, and remediation services	4.52%	3.62%	4.01%	3.31%
Finance and insurance	3.01%	3.05%	4.48%	3.23%
Transportation and warehousing	1.98%	2.77%	3.68%	3.03%
Wholesale trade	5.93%	3.34%	2.04%	2.42%
Information	2.19%	2.42%	1.99%	1.94%
Other services, except government	0.85%	0.58%	0.56%	0.54%
Health care and social assistance	0.80%	0.90%	0.66%	0.53%
Educational services	0.52%	0.63%	0.66%	0.52%
Accommodation and food services	0.28%	0.33%	0.37%	0.34%
Utilities	0.34%	0.32%	0.30%	0.33%
Retail trade	0.84%	0.60%	0.26%	0.10%
Real estate rental and leasing	0.31%	0.17%	0.09%	0.08%
Mining	0.59%	0.06%	0.06%	0.05%
Arts, entertainment, and recreation	0.03%	0.01%	0.02%	0.04%
Public administration	0.95%	0.32%	0.16%	0.03%
Agriculture, forestry, fishing, and hunting	0.02%	0.03%	0.01%	0.01%
Management of companies and enterprises	0.00%	0.00%	0.00%	0.00%

Notes: The table shows the share of total DoD procurement expenditure allocated to each sector categorized by the NAICS 2-digit code for the years 2005, 2010, 2015, and 2020.

The table above shows that the expenditure is concentrated in a few key sectors. The *manufacturing* sector is the largest recipient, accounting for nearly half of the total expenditure over time. *The professional, scientific, and technical services* (PSTS) sector also represents a significant share of the expenditure, receiving about one-quarter of the total, followed by the *construction* sector. Together, these three sectors comprise approximately 80% of the total DoD procurement expenditure.

Figure 3.9 illustrates the evolution of expenditure over time for the top three NAICS 2-digit sectors, along with the aggregated expenditure of the remaining 18 sectors.

Figure 3.9: Sectoral DoD procurement expenditure by NAICS 2-digit sectors (2001 - 2021)



Notes: *The figure shows the changes in DoD procurement spending across three primary sectors defined by NAICS 2-digit codes: manufacturing, professional scientific and technical services, and construction, as well as the consolidation of other sectors from 2001 to 2021. All values have been adjusted for inflation using the national defence government expenditure implicit price deflator and are presented in December 2021 dollars.*

The figure above supports the information from the table, indicating that the *manufacturing* sector receives the majority of expenditure, obtaining approximately twice as much as the second sector, *PSTS*, in most years. Expenditure in the *manufacturing* sector displays a positive trend over time and higher volatility, whereas the other sectors either fluctuate around a constant value or exhibit a very weak trend. Consequently, the share of military procurement expenditure allocated to the *manufacturing* sector has been increasing over time, as indicated in Table 3.1.

Table 3.2 analyses expenditure in the manufacturing sector at a more detailed level, examining its components using the 3-digit NAICS codes. Almost 70% of the expenditure in this sector is allocated to the subcategory of *transportation equipment manufacturing.* This expenditure is substantial as it encompasses the production of a variety of vehicles, from boats and aircraft to guided missiles. When looking at the 6-digit disaggregation, the main elements of this subcategory include *aircraft manufacturing*, *shipbuilding and repairing*, as well as *guided missiles* and *space vehicle manufacturing.*

Table 3.2: Total DoD procurement expenditure by manufacturing NAICS 3-digit categories

Manufacturing sector by NAICS 3-digit	**Total expenditure (Millions of dollars)**			
	2005	*2010*	*2015*	*2020*
Transportation equipment manufacturing	60,729.00	121,173.30	92,084.12	136,736.83
Computer and electronic product manufacturing	20,761.16	31,275.42	20,014.90	30,569.29
Chemical manufacturing	898.15	2,144.68	4,244.12	13,710.47
Fabricated metal product manufacturing	6,907.72	9,630.39	7,184.13	10,973.89
Machinery manufacturing	4,100.26	6,950.93	3,602.82	5,046.76
Petroleum and coal products manufacturing	6,241.31	5,785.83	5,649.57	3,709.05
Miscellaneous manufacturing	1,276.60	2,263.42	1,866.19	2,851.43
Food manufacturing	461.63	3,303.55	930.66	1,962.47
Apparel manufacturing	1,458.67	3,048.69	1,300.07	1,753.25
Electrical equipment, appliance, and component manufacturing	1,431.06	2,214.05	1,922.68	1,405.97
Textile product mills	919.63	573.99	178.09	536.46
Furniture and related product manufacturing	678.90	1,048.11	438.13	462.59
Plastics and rubber products manufacturing	281.69	384.09	469.85	217.37
Textile mills	259.47	88.42	88.83	164.13
Leather and allied product manufacturing	296.58	232.77	260.79	121.14
Primary metal manufacturing	276.79	242.46	162.09	118.98
Paper manufacturing	41.78	178.39	102.58	114.99
Wood product manufacturing	138.54	227.39	69.64	94.29
Nonmetallic mineral products	118.00	149.47	49.18	63.30
Printing and related support activities	516.26	24.53	26.21	26.24
Beverage and tobacco product manufacturing	276.58	390.74	12.22	8.11
Total	**108,069.78**	**191,330.60**	**140,656.85**	**210,647.01**

Notes: The table shows the total DoD procurement expenditure for the main subsectors classified by NAICS 3-digit code in the manufacturing sector for the years 2005, 2010, 2015, and 2020. The last row of the table shows the total expenditure in the manufacturing sector. The values presented are in millions of dollars and have been adjusted for inflation using the national defence government expenditure implicit price deflator. All amounts are expressed in December 2021 dollars.

Table 3.3 performs a similar analysis to the previous table, but illustrates the decomposition for the PSTS sector. The distribution of expenditure is more evenly distributed within this sector compared to the manufacturing sector. The three main subcategories are *architectural, engineering, and related services*; *scientific research and development services*; and *computer systems design and related services.*

Table 3.3: Total DoD procurement expenditure by PSTS NAICS 3-digit categories

PSTS sector by NAICS 3-digit	**Total expenditure (Millions of dollars)**			
	2005	*2010*	*2015*	*2020*
Architectural, engineering, and related Services	24,565.51	39,038.72	26,734.97	32,537.37
Scientific research and development services	21,779.67	32,492.45	24,084.80	32,024.62
Computer systems design and related services	12,173.37	13,508.93	13,810.38	20,666.00
Management, scientific, and technical consulting services	5,338.16	7,937.21	5,115.25	4,965.22
Other professional, scientific, and technical services	3,712.03	6,311.50	4,110.45	3,650.52
Accounting, tax preparation, bookkeeping, and payroll services	126.91	223.96	449.48	956.00
Advertising, public relations, and related services	691.17	654.08	687.77	812.47
Legal services	39.56	25.27	18.88	22.69
Specialized design services	629.11	53.16	34.78	20.93
Total	**69,055.50**	**100,245.28**	**75,046.74**	**95,655.82**

Notes: The table shows the total DoD procurement expenditure for the main subsectors classified by NAICS 3-digit code in the professional, scientific, and technical services (PSTS) sector for the years 2005, 2010, 2015, and 2020. The last row of the table shows the total expenditure in the PSTS sector. The values presented are in millions of dollars and have been adjusted for inflation using the national defence government expenditure implicit price deflator. All amounts are expressed in December 2021 dollars.

Table 3.4 complements the analysis by utilising the NAICS codes to examine the most disaggregated components of the classification system (6-digit). It displays the top 10 subcategories in overall expenditure for 2005, 2010, 2015, and 2020, including both the expenditure and the share of the total. This approach facilitates the identification of more specific items purchased by the DoD. *Aircraft manufacturing* consistently receives the highest expenditure share in most years presented in the table, often exceeding 10% of the total procurement expenditure. *Engineering services* also account for a significant share of the total expenditure. Additionally, items related to the production of all types of combat vehicles and missiles rank among the top categories in terms of expenditure. *Commercial and institutional building construction* represents a subcategory outside of manufacturing and services that receive considerable expenditure, with a share that fluctuates around 4% for 2010, 2015, and 2020.

Table 3.4: Top NAICS 6-digit industries by DoD procurement expenditure

Industry Title by NAICS 6-digit	Expenditure (Millions of Dollars)	Share of total expenditure
2005		
Engineering services	23,738.40	7.88%
Aircraft manufacturing	21,609.26	7.17%
Other aircraft parts and auxiliary equipment manufacturing	9,441.94	3.13%
Search, detection, navigation, guidance, aeronautical, and nautical system and instrument manufacturing	8,103.92	2.69%
Direct health and medical insurance carriers	7,251.52	2.41%
Military armored vehicle, tank, and tank component manufacturing	6,247.65	2.07%
Petroleum refineries	6,157.04	2.04%
Ship building and repairing	5,958.88	1.98%
Aircraft engine and engine parts manufacturing	5,929.88	1.97%
Other computer related services	5,286.64	1.75%
2010		
Aircraft manufacturing	44,887.26	11.10%
Engineering services	38,010.39	9.40%
Commercial and institutional building construction	19,280.29	4.77%
Guided missile and space vehicle manufacturing	15,711.59	3.88%
Ship building and repairing	15,378.92	3.80%
Other aircraft parts and auxiliary equipment manufacturing	13,594.67	3.36%
Military armored vehicle, tank, and tank component manufacturing	13,074.89	3.23%
Search, detection, navigation, guidance, aeronautical, and nautical system and instrument manufacturing	12,225.75	3.02%
Direct health and medical insurance carriers	12,086.88	2.99%
Other heavy and civil engineering construction	8,734.89	2.16%
2015		
Aircraft manufacturing	37,977.59	13.05%
Engineering services	26,070.19	8.96%
Ship building and repairing	15,505.61	5.33%
Other aircraft parts and auxiliary equipment manufacturing	13,463.77	4.63%
Direct health and medical insurance carriers	12,893.10	4.43%
Commercial and institutional building construction	10,707.07	3.68%
Guided missile and space vehicle manufacturing	10,030.22	3.45%
Search, detection, navigation, guidance, aeronautical, and nautical system and instrument manufacturing	10,002.86	3.44%
Petroleum refineries	5,619.62	1.93%
Other support activities for air transportation	5,401.85	1.86%
2020		
Aircraft manufacturing	53,047.46	13.03%
Engineering services	31,600.54	7.76%
Ship building and repairing	25,156.95	6.18%
Guided missile and space vehicle manufacturing	23,729.21	5.83%
Commercial and institutional building construction	16,739.81	4.11%
Research and development in the physical, engineering, and life sciences (except nanotechnology and biotechnology)	16,644.61	4.09%
Search, detection, navigation, guidance, aeronautical, and nautical system and instrument manufacturing	16,518.13	4.06%
Other aircraft parts and auxiliary equipment manufacturing	15,521.46	3.81%
Direct health and medical insurance carriers	12,928.44	3.18%
Aircraft engine and engine parts manufacturing	8,533.76	2.10%

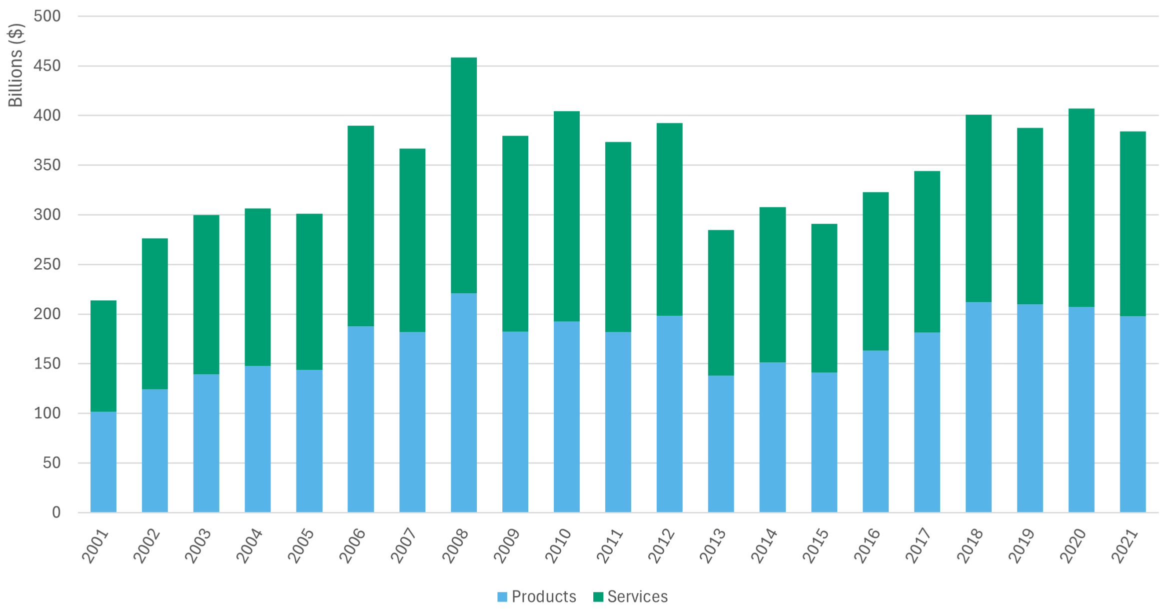
Notes: The table presents the top 20 industries classified by the NAICS 6-digit code based on total DoD procurement expenditures for the years 2005, 2010, 2015, and 2020. The second column lists the total expenditures in millions of dollars, while the third column indicates the percentage share of the total expenditure for each industry. The values presented are in millions of dollars and have been adjusted for inflation using the national defence government expenditure implicit price deflator. All amounts are expressed in December 2021 dollars.

The tables in this subsection provide evidence for a sectoral bias in military government procurement expenditure, as highlighted by Cox et al. (2024). This expenditure is not only concentrated in a limited number of sectors, but these sectors also display a significantly smaller proportion of private expenditure. In essence, the distribution of procurement military expenditure differs markedly from that of private expenditure.

To complement the sectoral analysis, I use Product and Service Codes (PSC) to classify DoD procurement expenditure. The PSC classification enables a clearer distinction between goods and services, and further allows services to be separated into research and development (R&D) and non-R&D categories. Figure 3.10 displays the total expenditure for each year in the data and indicates the allocation between products and services.

As can be observed from the figure, the total expenditure is consistently evenly distributed between the two categories over time. Notably, the distribution remained close even in years of significant increases and decreases in overall expenditure, such as in 2008 and 2013, respectively. The year with the most uneven distribution between the two categories is 2002, in which products accounted for 45% of the total expenditure while services accounted for 55%.

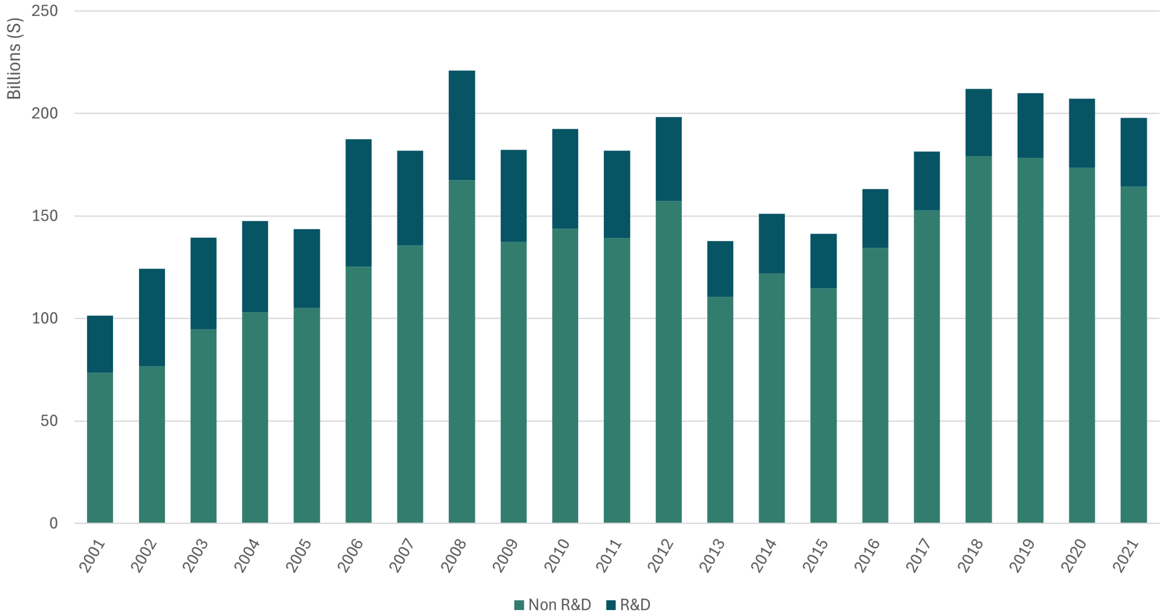
Figure 3.10: Products and services DoD procurement expenditure by PSC codes



Notes: *The figure displays the yearly DoD procurement expenditure from 2001 to 2021, distinguishing between the products and services sectors. This differentiation was made using Product and Service Codes (PSC). All values have been adjusted for inflation using the national defence government expenditure implicit price deflator and are presented in December 2021 dollars.*

Figure 3.11 illustrates the proportion of expenditure on services that is dedicated to R&D. It shows a significant decline in the share of services expenditure allocated to R&D over time. Initially, R&D represented approximately 30% of the total services expenditure, but this percentage decreased to around 15% by the end of the data period.

Figure 3.11: R&D and non-R&D services expenditure by PSC codes



Notes: *The figure illustrates the annual DoD procurement expenditure for the services sector from 2001 to 2021, distinguishing between research and development (R&D) and non-R&D expenditures. This differentiation is based on Product and Service Codes (PSC). All values have been adjusted for inflation using the national defence government expenditure implicit price deflator and are presented in December 2021 dollars.*

The figure above also indicates a substantial decrease in services expenditure in 2013, from which the R&D segment did not fully recover by 2021. The maximum expenditure on R&D was observed in 2006, with a value of approximately $62.1 billion. In 2012, this value was $41.0 billion, and as previously mentioned, it did not recover to the previous levels in the analysed data frame, reaching $33.5 billion in 2021. However, the same cannot be said for the non-R&D segment of the services sector, which achieved its highest value in 2018 ($179.1 billion), maintaining expenditure levels comparable to 2008 from 2018 onwards.

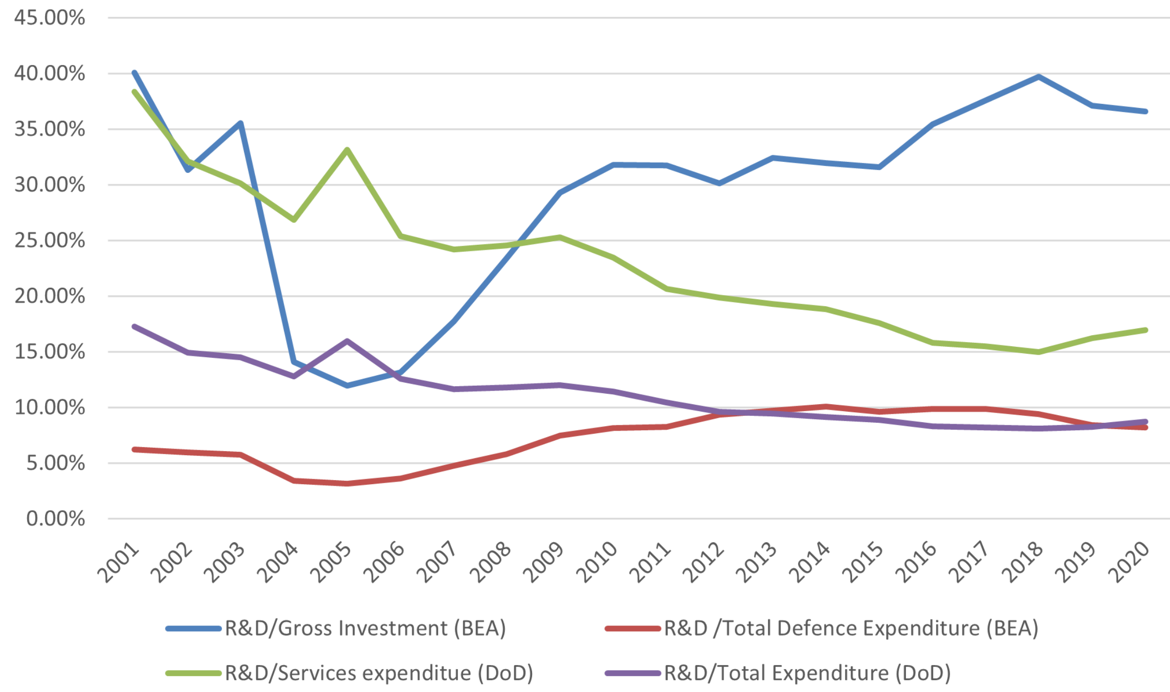
When R&D expenditure is expressed as a share of total DoD procurement expenditure, it displays a similar downward trend. This pattern should be noted because it moves in the opposite direction from the defence R&D shares reported in the national accounts by the BEA[10] .

The figure below compares four measures of the R&D share. The first two are cal-

[10]Data available at `https://fred.stlouisfed.org/release/tables?rid=53&eid=16936`.

culated from the procurement data used in this chapter, using services expenditure (as in Figure 3.11) and total DoD expenditure as denominators, respectively. The other two are constructed from BEA Table 3.11.5. The first BEA measure is defence R&D expenditure divided by total defence expenditure, measured as national defence consumption expenditures plus gross investment. The second is defence R&D expenditure divided by defence gross investment. These two BEA-based measures are reported because they capture different margins: the former shows the share of overall defence spending accounted by R&D, while the latter shows the share of defence investment accounted by R&D.

Figure 3.12: Evolution of defence R&D expenditure shares: national accounts and DoD procurement data.



Notes: *The figure above displays the evolution of different measurements of the share of defence expenditure allocated to research and development (R&D). The blue and red lines represent the share of R&D expenditure in the national accounts relative to gross investment and total defence expenditure, respectively, using data sourced from the BEA. The green and purple lines represent the corresponding shares calculated from the procurement data used in this chapter: the former is the ratio of R&D expenditure to services expenditure, and the latter is the ratio of R&D expenditure to total expenditure, with both measures classified using PSC codes.*

The figure highlights the contrast clearly. While the R&D shares constructed from the procurement data presented in this chapter display a clear downward trend, the BEA-based measures suggest an upward trend in the share of defence expenditure allocated to

R&D, particularly when R&D is scaled by total defence expenditure.

These patterns differ because the underlying measures are conceptually distinct. The procurement-based data used in this chapter capture only expenditure associated with R&D service contracts identified through PSC codes. By contrast, the BEA series is a national-accounts measure of defence R&D expenditure and includes several components beyond procurement contracts, such as grants and own-account government R&D recorded as intellectual property investment.

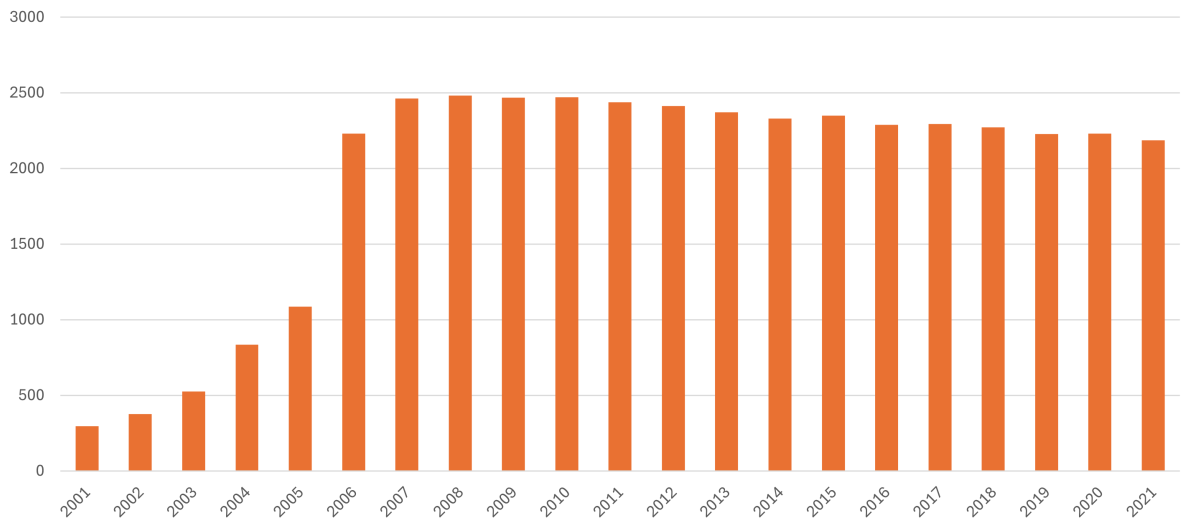
The decline in the procurement-based shares can therefore be interpreted as a shift in the composition of DoD expenditure within the procurement data. Over time, R&D-related service contracts account for a smaller share of both services expenditure and total procurement expenditure. This partly reflects the sharp reduction in R&D expenditure observed around 2012 and its incomplete recovery thereafter, while other procurement categories expanded more strongly. By contrast, the increase in the BEA-based shares suggests that defence R&D gained relative importance within the broader national-accounts framework of defence expenditure, even if this broader concept is not directly comparable with the procurement-based measure used in this chapter.

To conclude this section, it is important to acknowledge a key limitation of sectoral aggregation using either NAICS or PSC classification systems: prime contractors often subcontract portions of the work, particularly in high-value procurements. As a result, aggregating data by the NAICS or PSC code of the prime award can misrepresent the actual distribution of economic activity across sectors. For example, a $50 million contract for aircraft or submarine production may involve substantial subcontracts in areas such as software development or engineering services. These secondary activities are not reflected in the prime contract's NAICS classification, leading to potential overstatement of certain sectors and understatement of others.

3.3.5 Geographical distribution

This section examines the geographical distribution of the military procurement expenditure at the county level. Figure 3.13 begins the analysis by illustrating how many of the 3,148 counties had at least one procurement awarded each year, excluding U.S. overseas territories. From 2001 to 2021, a total of 2,942 distinct counties received at least one procurement DoD contract. As illustrated in the figure, the number of counties receiving procurement expenditures has increased significantly since 2006. This increase arises from enhancements in the data's reporting standards. Prior to 2005, most procurement registered did not specify the place of performance location, leading to difficulties in associating those contracts with the corresponding counties. Therefore, any regionally focused analysis using these data should be mindful of this feature when working with observations from 2005 and earlier.

Figure 3.13: Annual number of counties with awarded DoD procurements



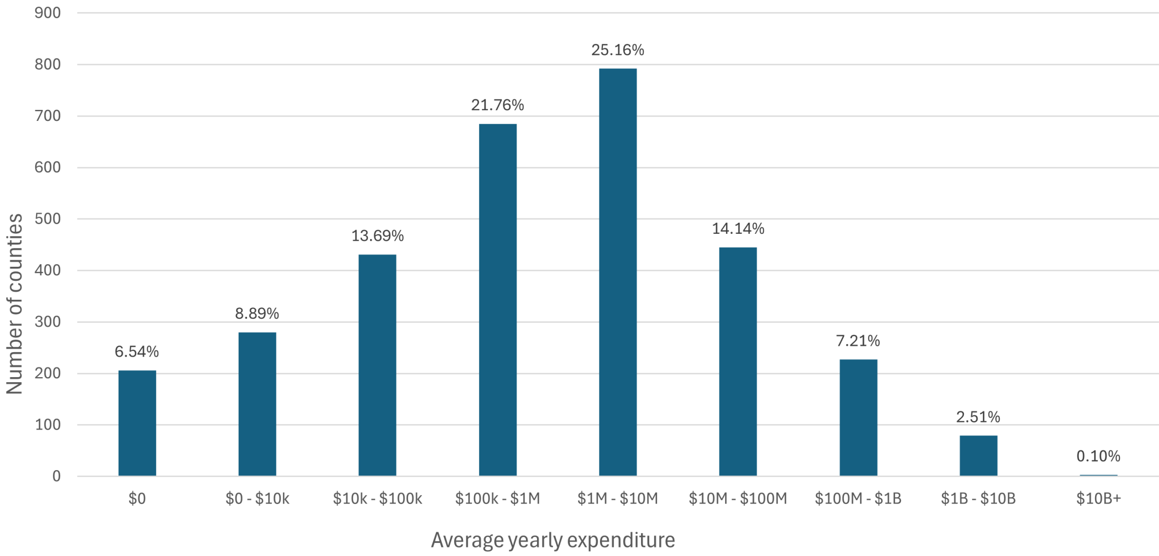
Notes: *The figure illustrates the total number of counties awarded at least one DoD procurement each year from 2001 to 2021.*

Figure 3.14 shows the distribution of expenditure across all counties. Each county was categorized into a spending group based on its average yearly procurement expenditure[11]. The figure reveals that many counties received low expenditure levels, with 917 counties (29.1%) receiving yearly average values of up to $100,000. Most counties received average expenditure ranging from $100,000 to $100 million. Approximately 10% of the counties received significant amounts of military procurement expenditures. Specifically,

[11]The average values presented in the figure is from 2006 to 2021, excluding the initial years as previously discussed.

227 counties (7.2%) reported yearly averages between \$100 million and \$1 billion, while 79 counties (2.5%) averaged between \$1 billion and \$10 billion. Additionally, 3 counties (0.1%) received average amounts exceeding \$10 billion.

Figure 3.14: Distribution of counties by average yearly DoD procurement expenditure



Notes: *The figure illustrates the distribution of counties based on their average yearly DoD procurement expenditures from 2001 to 2021. The counties have been categorized into different groups according to their average DoD procurement expenditure values to represent the distribution. The percentages displayed at the top of each bar represent the share of total counties included in that category. All figures have been adjusted for inflation using the national defence government expenditure implicit price deflator and are expressed in December 2021 dollars.*

Table 3.5 complements the information presented in the previous figure. It details the top 20 counties in terms of expenditure for 2010, 2015, and 2020. The table displays the expenditure levels for those years along with the share of total expenditure. It is noticeable that some counties account for a significant share of the total expenditure. Furthermore, despite a few changes from year to year, the group of counties comprising the top 20 in expenditure each year remains largely the same, indicating that the government frequently renews existing high-value contracts with companies located in these counties.

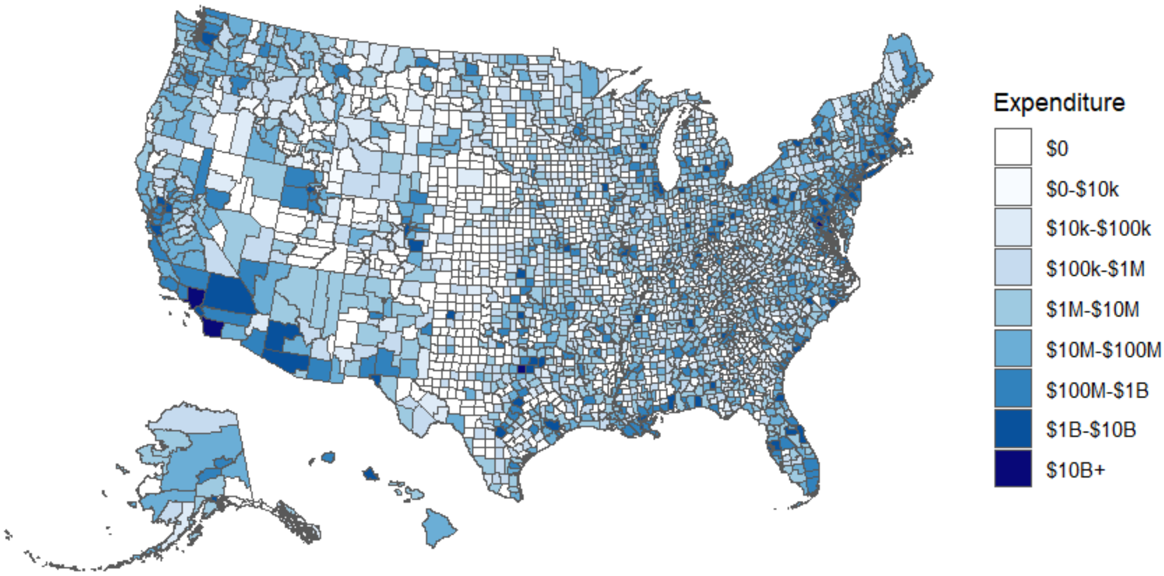
Table 3.5: Top 20 counties by DoD procurement expenditure for 2010, 2015, and 2020

2010		2015		2020	
County	**Total expenditure (millions of dollars) and share of total**	**County**	**Total expenditure (millions of dollars) and share of total**	**County**	**Total expenditure (millions of dollars) and share of total**
Fairfax, VA	19,653.51 (4.89%)	*Fairfax, VA*	15,123.48 (5.22%)	*Tarrant, TX*	26,583.70 (6.56%)
San Diego, CA	14,056.55 (3.50%)	*Tarrant, TX*	11,762.80 (4.06%)	*Fairfax, VA*	16,444.80 (4.06%)
Tarrant, TX	13,604.80 (3.39%)	*San Diego, CA*	9,684.60 (3.34%)	*Dallas, TX*	13,924.45 (3.44%)
Los Angeles, CA	12,670.95 (3.15%)	*Madison, AL*	6,732.59 (2.32%)	*Middlesex, MA*	12,879.70 (3.18%)
Winnebago, WI	8,536.48 (2.12%)	*St. Louis, MO*	6,125.91 (2.12%)	*St. Louis, MO*	12,537.75 (3.09%)
Santa Clara, CA	7,150.61 (1.78%)	*Los Angeles, CA*	6,080.66 (2.10%)	*San Diego, CA*	11,528.82 (2.85%)
Middlesex, MA	7,065.33 (1.76%)	*New London, CT*	6,034.78 (2.08%)	*New London, CT*	9,930.59 (2.45%)
Maricopa, AZ	6,594.28 (1.64%)	*Pima, AZ*	5,084.53 (1.76%)	*Pima, AZ*	9,464.70 (2.34%)
Madison, AL	6,398.92 (1.59%)	*Middlesex, MA*	5,074.42 (1.75%)	*Los Angeles, CA*	7,753.15 (1.91%)
Cobb, GA	6,273.35 (1.56%)	*Dallas, TX*	4,998.05 (1.73%)	*Jefferson, KY*	7,626.81 (1.88%)
Essex, MA	5,933.91 (1.48%)	*Orange, FL*	4,788.07 (1.65%)	*Madison, AL*	7,578.60 (1.87%)
Arlington, VA	5,868.67 (1.46%)	*Jefferson, KY*	4,506.02 (1.56%)	*District of Columbia, DC*	6,888.71 (1.70%)
Pima, AZ	5,822.48 (1.45%)	*King, WA*	4,502.23 (1.55%)	*Orange, FL*	6,298.50 (1.55%)
District of Columbia, DC	5,634.13 (1.40%)	*District of Columbia, DC*	4,250.71 (1.47%)	*Fairfield, CT*	5,124.89 (1.26%)
Jefferson, KY	5,000.29 (1.24%)	*Santa Clara, CA*	3,908.12 (1.35%)	*Hartford, CT*	4,948.13 (1.22%)
St. Louis, MO	4,975.78 (1.24%)	*Cobb, GA*	3,849.45 (1.33%)	*King, WA*	4,802.77 (1.19%)
St. Louis City, MO	4,846.19 (1.21%)	*Allegheny, PA*	3,796.98 (1.31%)	*Virginia Beach City, VA*	4,544.40 (1.12%)
Fairfield, CT	4,811.50 (1.20%)	*Hennepin, MN*	3,702.84 (1.28%)	*Jackson, MS*	4,421.71 (1.09%)
Hartford, CT	4,665.23 (1.16%)	*Fairfield, CT*	3,689.37 (1.27%)	*Anne Arundel, MD*	4,236.60 (1.05%)
New London, CT	4,547.85 (1.13%)	*Sacramento, CA*	3,682.39 (1.27%)	*Newport News City, VA*	4,174.90 (1.03%)

Notes: The table displays the top 20 counties ranked by DoD procurement expenditures for the years 2010, 2015, and 2020. It includes the corresponding expenditures in millions of dollars, as well as each county's share of the total expenditure (shown in parentheses). All values have been adjusted for inflation using the national defence government expenditure implicit price deflator and are expressed in December 2021 dollars.

Figure 3.15 presents a visual analysis of the geographical distribution of expenditure by displaying a map that illustrates spending levels across each county in 2010. The map reveals a concentration of expenditure in the states of California, Connecticut and Virginia. Furthermore, a substantial level of expenditure is found in the state of Texas, particularly in Tarrant County and its neighbouring counties.

Figure 3.15: County-level DoD procurement expenditure (2010)

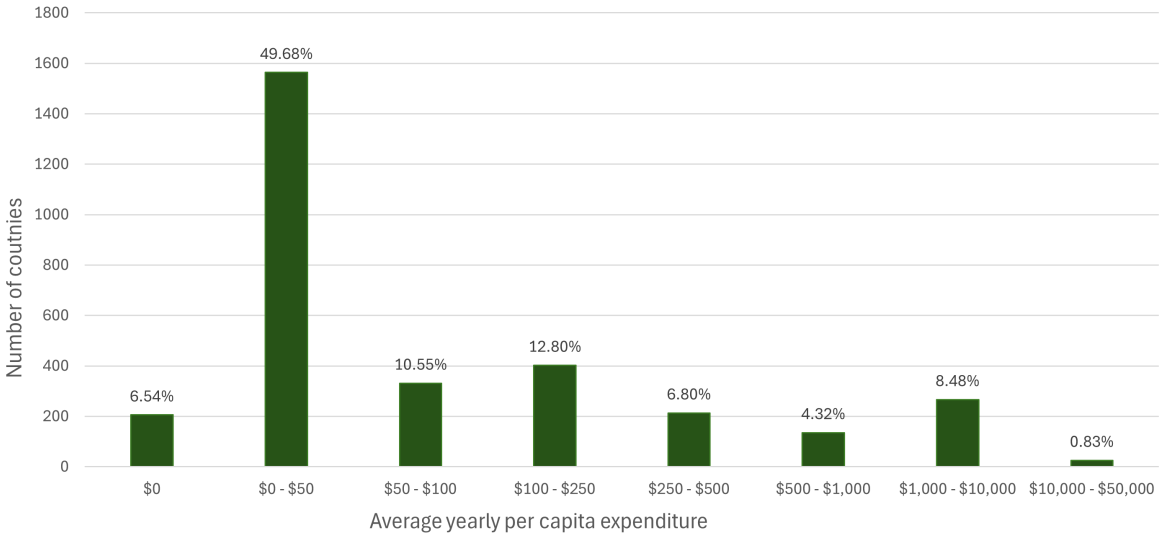


Notes: *The map displays the geographical distribution of the county-level DoD procurement expenditure for 2010. All values have been adjusted for inflation using the national defence government expenditure implicit price deflator and are presented in December 2021 dollars.*

Appendix B presents a set of figures and tables similar to Table 3.5 and Figure 3.15 but examines the geographical distribution of expenditure for the two main sectors according to the NAICS 2-digit classification. Table B1 and Figure B1 illustrate the county-level expenditure distribution of the manufacturing sector, while Table B2 and Figure B2 present the distribution for the PSTS sector. The figures indicate a similar spatial distribution for both sectors; however, in many instances, one sector predominates over the other, as can be observed with the aid of the tables. For instance, in the state of Virginia, the services sector predominates, whereas in Texas, expenditure in manufacturing is higher.

It is important to note that many of the counties listed in Table 3.5 have relatively large populations, which may suggest that more populous counties received higher levels of expenditure. To account for population size, the following figures and table reanalyse the data, presenting military procurement expenditure on a per capita basis. Figure 3.16 illustrates the distribution of the average annual per capita expenditure.

Figure 3.16: Distribution of counties by per capita average yearly DoD procurement expenditure



Notes: *The figure illustrates the distribution of counties based on their average yearly per capita DoD procurement expenditures from 2001 to 2021. The counties have been categorized into different groups according to their average DoD procurement expenditure values to represent the distribution. The percentages displayed at the top of each bar represent the share of total counties included in that category. All figures have been adjusted for inflation using the national defence government expenditure implicit price deflator and are expressed in December 2021 dollars.*

As the figure illustrates, 206 counties received no expenditure, whilst another 1,564 counties experienced an average annual per capita value of less than $50, accounting for 56.2% of all counties. Additionally, 1,085 counties (34.5%) had intermediate average per capita expenditures, ranging from $50 to $1,000. A total of 267 counties received a considerably high amount, ranging from an average of $1,000 to $10,000, while a group of 26 counties had an even more significant amount, ranging from $10,000 to $50,000.

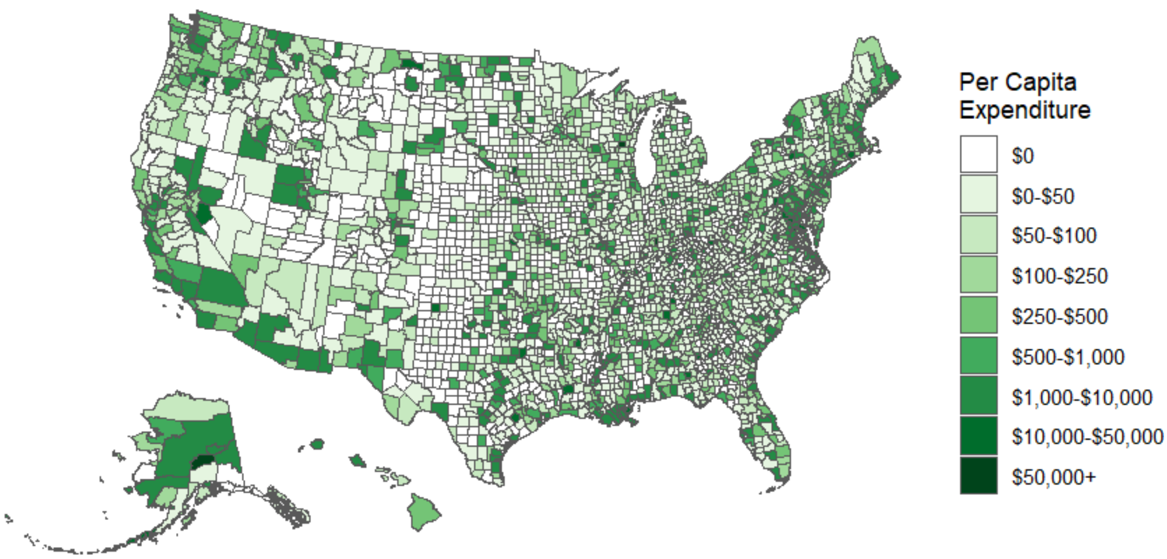
Table 3.6 presents the top 20 counties ranked by average per capita expenditure for 2010, 2015, and 2020. Notably, the counties in this new table differ significantly from those in Table 3.5, which lists the top counties in terms of total average expenditure. Figure 3.17 complements the analysis by providing a visual representation of the geographical distribution of per capita expenditure for 2010.

Table 3.6: Top 20 counties by per capita DoD procurement expenditure for 2010, 2015, and 2020

2010		2015		2020	
County	**Per capita expenditure**	**County**	**Per capita expenditure**	**County**	**Per capita expenditure**
Martin, IN	66,867.02	*Pulaski, IL*	45,262.66	*Tioga, NY*	50,868.89
Denali, AK	57,873.89	*Denali, AK*	43,159.15	*Sagadahoc, ME*	48,434.10
Winnebago, WI	51,072.63	*Sagadahoc, ME*	30,913.78	*Denali, AK*	48,303.34
Falls Church City, VA	50,654.64	*Martin, IN*	30,108.92	*Martin, IN*	45,397.37
Colorado, TX	46,155.23	*Calhoun, AR*	28,303.74	*New London, CT*	36,992.33
Chattahoochee, GA	40,581.37	*New London, CT*	22,366.11	*Queen Anne's, MD*	31,281.64
Pulaski, IL	37,116.27	*Mineral, NV*	20,780.18	*Jackson, MS*	30,840.63
Calhoun, AR	32,984.66	*Tioga, NY*	19,851.82	*Summers, WV*	28,866.73
King George, VA	28,589.36	*Chattahoochee, GA*	19,501.72	*Chattahoochee, GA*	24,071.29
Arlington, VA	28,036.97	*Potter, TX*	19,402.57	*Atchison, MO*	23,040.90
Potter, TX	26,424.97	*King George, VA*	19,289.60	*Newport News City, VA*	22,440.04
St. Mary's, MD	23,710.20	*Madison, AL*	19,062.02	*King George, VA*	21,762.46
Union, SD	23,279.78	*Dale, AL*	18,010.28	*St. Mary's, MD*	21,278.96
Sagadahoc, ME	23,090.69	*Hunt, TX*	17,677.87	*Marinette, WI*	20,658.35
Hunt, TX	22,517.28	*St. Mary's, MD*	16,620.71	*Madison, AL*	19,407.27
Dale, AL	20,564.47	*Jackson, MS*	15,821.12	*Dale, AL*	19,380.09
Mineral, NV	20,548.70	*Newport News City, VA*	14,266.25	*Hunt, TX*	19,338.02
Newport News City, VA	20,319.88	*Klickitat, WA*	14,058.15	*Luna, NM*	17,049.01
Manassas City, VA	19,922.39	*Fairfax, VA*	13,254.12	*Potter, TX*	16,558.45
Madison, AL	19,037.94	*Yolo, CA*	13,053.66	*Plaquemines, LA*	16,140.81

Notes: The table displays the top 20 counties ranked by per capita DoD procurement expenditures for the years 2010, 2015, and 2020, including the corresponding expenditure. All values have been adjusted for inflation using the national defence government expenditure implicit price deflator and are expressed in December 2021 dollars.

Figure 3.17: County-level per capita DoD procurement expenditure (2010)



Notes: *The map displays the geographical distribution of the county-level per capita DoD procurement expenditure for 2010. All values have been adjusted for inflation using the national defence government expenditure implicit price deflator and are presented in December 2021 dollars.*

By observing the figure above, some states exhibit a few clusters of counties with high expenditure per capita, including Alaska, Arizona, California, and Utah. The most prominent are the states of Maryland and Virginia, which encompass a substantial number of nearby counties with elevated per capita expenditure. While the patterns are analogous to those in Figure 3.15, the per capita values particularly highlight low-population states such as Alaska and Utah.

Geographical Herfindahl-Hirschman Index To conclude the geographical analysis of the DoD expenditure distribution, I present a geographical Herfindahl-Hirschman Index (HHI). This index is a concentration measurement defined as the sum of each each county's squared expenditure shares. Although it is commonly used in the context of firm competition and market concentration, the same logic can be applied to other settings. In this analysis, I use it to measure the geographical concentration of DoD procurement expenditure across counties. The index for year t is calculated as follows:

$$HHI_t = \sum_{i=1}^{N} (DoD_{i,t}^{share})^2, \tag{3.1}$$

in which N denotes the number of counties, and $DoD_{i,t}^{share}$ denotes the share of the DoD procurement expenditure allocated to county i in the year t.

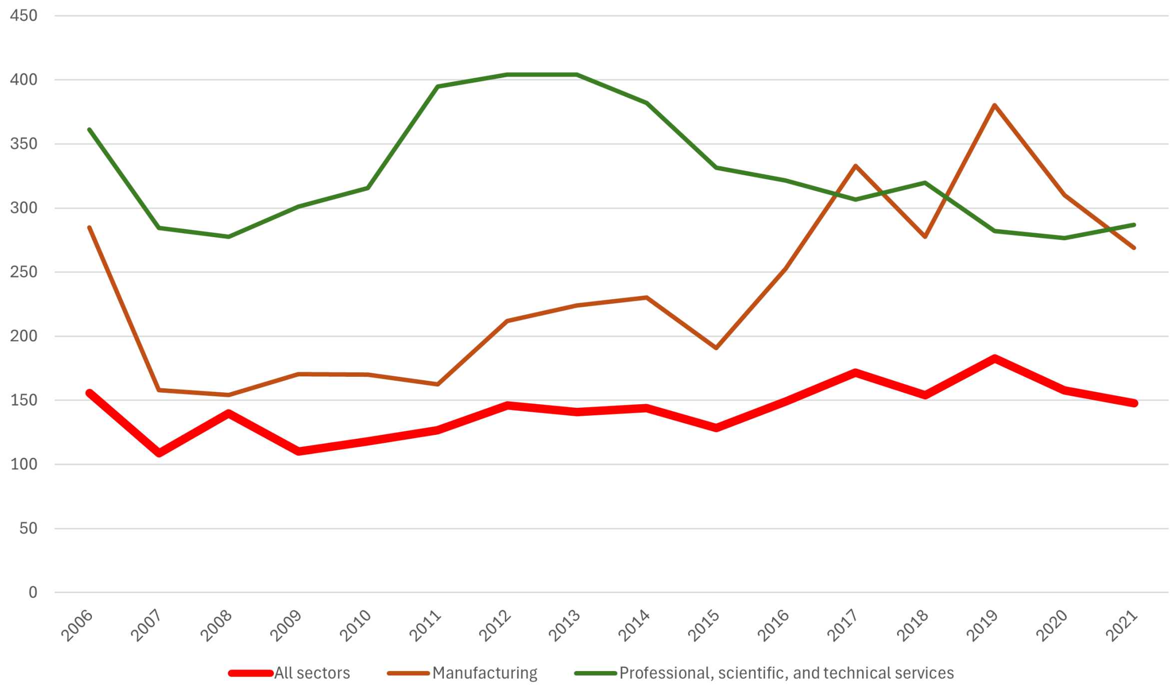
Since the shares are expressed in percentage points, the index ranges from $\frac{10{,}000}{N}$, when expenditure is evenly distributed across all counties, to 10,000, when all expenditure is concentrated in a single county[12]. Higher values of the index, therefore, indicate greater geographical concentration.

The HHI is useful because it accounts not only for the number of counties receiving expenditure, but also for the relative size of the amounts received by each county. Figure 3.18 plots the index from 2007 to 2021 for total procurement expenditure and for the two major NAICS 2-digit sectors: Manufacturing and PSTS[13].

[12]Given that the number of counties is finite, the lower bound is not zero. With 3,143 counties, an equal distribution across all counties would imply an HHI of approximately 3.18.

[13]The years prior to 2007 are excluded due to the reporting issues discussed earlier in this chapter. In those years, many contracts did not report the place of performance, which would artificially inflate the measured concentration of expenditure across counties.

Figure 3.18: Geographical HHI index



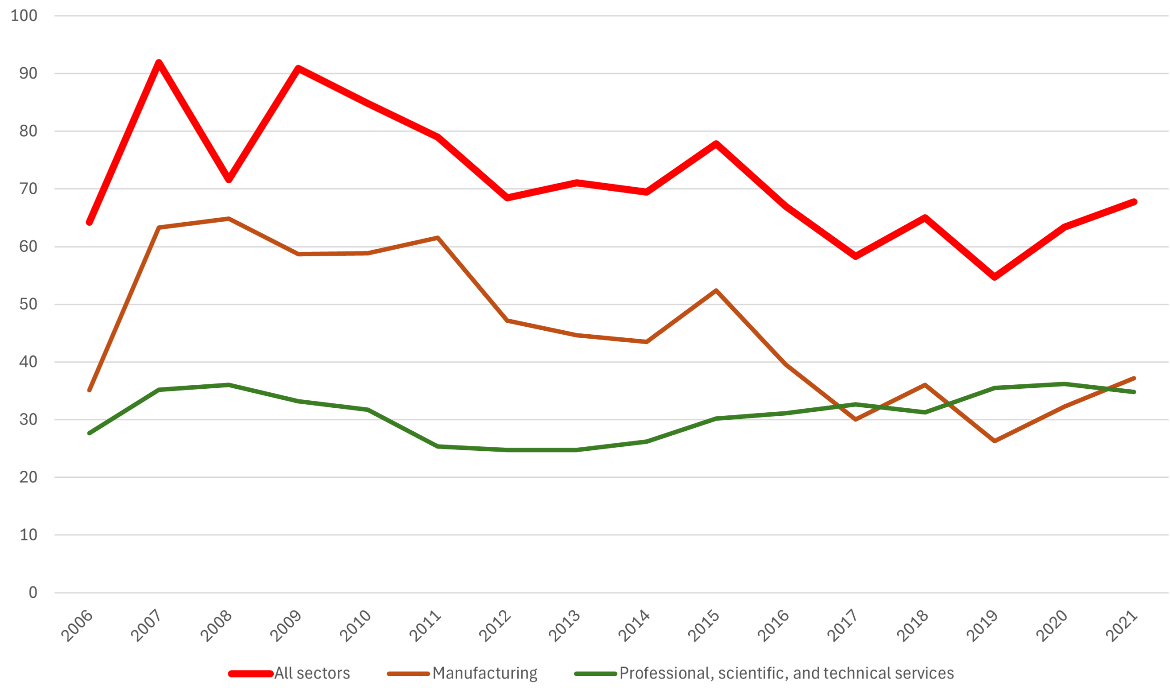
Notes: *The figure above displays the evolution of the county-level geographical HHI for the overall DoD expenditure and also for Manufacturing and Professional, scientific, and technical services sectors by NAICS 2-digit classification. The data range is 2006 to 2021.*

One intuitive way of analysing the HHI is by building a measurement called the effective number of counties (ENC), which is an application of what was proposed by Adelman (1969) to examine the concentration of firms. The ENC is defined as the inverse of the HHI multiplied by 10,000:

$$ENC_t = \frac{10,000}{HHI_t} \tag{3.2}$$

This measure indicates the number of equally sized counties, in terms of expenditure shares, that would generate the same level of geographical concentration observed in the data. For example, an HHI of 100 corresponds to an ENC of 100, meaning that the observed distribution is as concentrated as if expenditure were evenly distributed across 100 counties. The ENC for total expenditure, Manufacturing, and PSTS is shown in Figure 3.19.

Figure 3.19: Effective number of counties (ENC)



Notes: *The figure above displays the evolution of the effective number of counties (ENC) for the overall DoD expenditure and also for Manufacturing and Professional, scientific, and technical services sectors by NAICS 2-digit classification. The data range is 2006 to 2021.*

Figures 3.18 and 3.19 reveal a considerable degree of geographical concentration in DoD procurement expenditure, as well as important heterogeneity across sectors. For total expenditure, the HHI remains well above the equal-distribution benchmark and ranges from 108.8 in 2007 to a peak of 182.8 in 2019. The latter value corresponds to an ENC of 54.7, meaning that the distribution of total procurement in 2019 was as concentrated as if expenditure were evenly allocated across roughly 55 counties. This indicates that, although many counties receive some procurement expenditure, the effective geographical concentration is much narrower.

The sectoral patterns show an even stronger heterogeneity. Manufacturing displays a clear increase in geographical concentration over time. At the beginning of the period, its HHI is relatively close to the aggregate index, but it rises substantially in the late 2010s, reaching 380.3 in 2019. This corresponds to an ENC of 26.3, suggesting that Manufacturing procurement became highly concentrated in a small number of counties. PSTS, by contrast, is consistently more concentrated than total expenditure for most of the sample, but its dynamics differ from those of Manufacturing. Its HHI rises sharply up to around 2012 and 2013, before declining gradually and returning closer to its initial levels by 2021. This suggests that concentration in PSTS is persistently high, but does not display the same sustained upward trend observed in Manufacturing.

The results shown in Figures 3.18 and 3.19 complement the county-share evidence presented in Table 3.5 and Tables B1 and B2 in Appendix B. Table 3.5 indicates that the share of total expenditure directed to the top recipient counties increases over time. This pattern is even more pronounced in Manufacturing, where Table B1 shows that Tarrant County, TX, the top recipient county, increases its share from 5.77% in 2010 to 11.66% in 2020. By contrast, Table B2 does not suggest a comparable upward trend among the top PSTS counties, which is consistent with the behaviour of the sectoral HHI. Overall, the HHI and ENC measures reinforce the conclusion that DoD procurement is geographically concentrated, and that this concentration is both sector-specific and evolving over time.

3.3.6 Top companies in DoD procurement expenditure

This section adds to the analysis by showing the distribution of the DoD procurement expenditure among the top recipient companies. Table 3.7 presents the top companies by expenditure for 2010 and 2020. The table shows the amount allocated to each listed company each year, along with the proportion of the total expenditure it represents. The companies were identified using their UEI codes; however, the same company may be registered under different UEI codes. To link the expenditure to each company, its name was merged into the procurement data using its UEI code, and the aggregation was carried out based on the company name. In some instances, however, the company appears under different names, as in the case of *The Boeing Company*. The table lists the companies without altering their names, using those provided in the UEI codes list from `sam.gov`.

Table 3.7: Top 20 companies by DoD procurement expenditure for 2010 and 2020

Company name	Total expenditure (millions of dollars)	Share of total expenditure
	2010	
Lockheed Martin Corporation	35,203.82	8.70%
Raytheon Company	15,378.57	3.80%
Mcdonnell Douglas Corporation	10,714.57	2.65%
Northrop Grumman Systems Corporation	8,036.62	1.99%
Oshkosh Corporation	7,906.39	1.95%
The Boeing Company	6,344.43	1.57%
Electric Boat Corporation	4,436.51	1.10%
Humana Military Healthcare Services Incorporated	4,085.45	1.01%
Sikorsky Aircraft Corporation	4,030.13	1.00%
United Technologies Corporation	3,988.61	0.99%
L-3 Communications Corporation	3,744.82	0.93%
Health Net Federal Services, LLC	3,717.07	0.92%
Triwest Healthcare Alliance Corporation	3,608.74	0.89%
Bell Boeing Joint Project Office	3,443.30	0.85%
Huntington Ingalls Incorporated	3,339.38	0.83%
Science Applications Internati	3,290.80	0.81%
Booz Allen Hamilton Inc.	3,133.59	0.77%
Boeing Company, The	2,847.56	0.70%
General Electric Company	2,769.17	0.68%
Computer Sciences Corporation	2,616.48	0.65%
	2020	
Lockheed Martin Corporation	53,564.41	13.16%
Raytheon Company	17,509.28	4.30%
Northrop Grumman Systems Corporation	10,552.18	2.59%
Mcdonnell Douglas Corporation	10,045.34	2.47%
Electric Boat Corporation	9,554.25	2.35%
Huntington Ingalls Incorporated	7,782.25	1.91%
Boeing Company, The	7,535.79	1.85%
Humana Military Healthcare Services Incorporated	7,183.09	1.76%
The Boeing Company	5,069.24	1.25%
Sikorsky Aircraft Corporation	4,906.40	1.21%
United Technologies Corporation	4,636.75	1.14%
General Electric Company	3,855.77	0.95%
Atlantic Diving Supply, Inc.	3,428.26	0.84%
Modernatx, Inc.	3,375.49	0.83%
Health Net Federal Services, LLC	3,261.71	0.80%
L-3 Communications Corporation	3,092.52	0.76%
Mckesson Corporation	3,071.89	0.75%
Oshkosh Defense, LLC	2,594.77	0.64%
Booz Allen Hamilton Inc.	2,371.86	0.58%
Science Applications International Corporation	2,363.03	0.58%

Notes: The table presents the top 20 companies ranked by their DoD procurement expenditures for the years 2010 and 2020. The companies' names were identified using their Unique Entity Identification (UEI) codes, and the total expenditures were aggregated based on their registered names in the UEI dictionary. This may result in the same company appearing twice if it is registered under different names or with any discrepancies. The second column lists the total expenditures in millions of dollars, while the third column indicates the percentage share of the total expenditure for each company. The values presented are in millions of dollars and have been adjusted for inflation using the national defence government expenditure implicit price deflator. All amounts are expressed in December 2021 dollars.

As the table suggests, the majority of companies in the top expenditure ranking include aerospace, missile, technology, and vehicle manufacturers. Notable exceptions are health insurance companies such as *Humana Inc.* and *Health Net Federal Services*. It is evident that certain companies command a significant share of the total expenditure. For instance, *Lockheed Martin Corporation*, an aerospace manufacturer, accounts for approximately 10% of the total expenditure each year. The company engages in procurement-related activities in many counties, although the majority of these activities are based in *Tarrant County, Texas*. *Raytheon Company*, one of the world's largest guided missile producers, also receives a considerable share of the DoD procurement expenditure. Its headquarters is currently located in *Arlington, Virginia*, having previously been situated in Massachusetts, a location that garnered a significant portion of this company's procurement expenditure. However, a substantial amount of the production activities occur in *Collin County, Texas*, and *Pima County, Arizona*.

When analysing the information provided by Table 3.7, in conjunction with the county-level data from Table 3.5, we can conclude that some counties with high expenditure owe this feature to a single company located in that area. For example, *Tarrant County* received over $24 billion in total expenditure in 2020; however, as Table 3.7 indicates, nearly all of that expenditure originates from contracts awarded to *Lockheed Martin Corporation*. A similar situation occurs in other counties presented in Table 3.5, such as *Fairfax* and *St. Louis*, where one single company is accountable for the vast majority of the procurement expenditure.

3.3.7 Obligations and actual spending

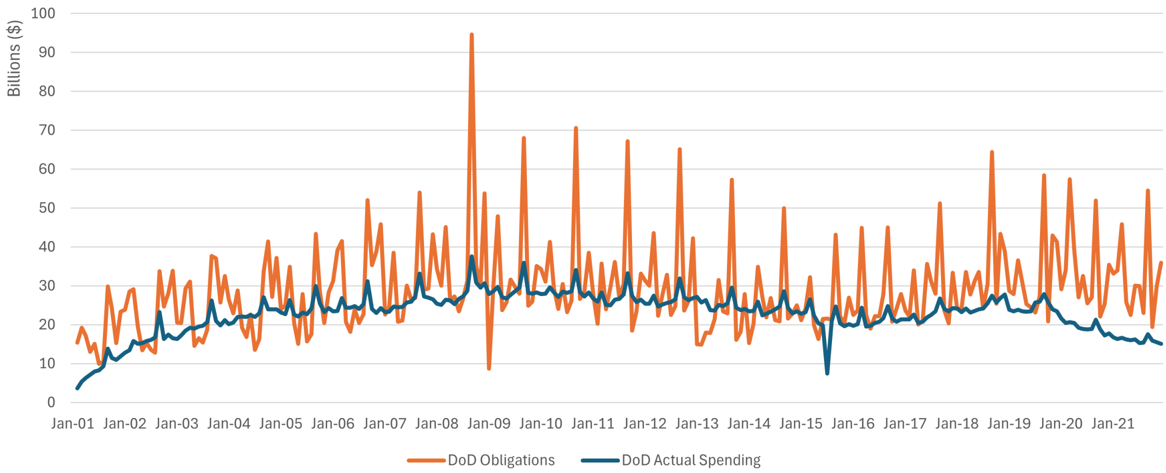
In contrast to the data presented so far in this section, this subsection discusses the behaviour of the aggregated expenditure data when contract lengths are considered during the aggregation process. Following Demyanyk et al. (2019), I refer to the total value of all procurements signed on a given date, regardless of duration, as DoD obligations. The second measurement of expenditure I present is termed DoD actual spending[14].

[14]This distinction is made only in this subsection. In all other sections of the chapter, the expenditure values provided refer to the DoD obligations, or the total amount allocated to contracts signed during

To calculate it, the expenditure is evenly distributed across all months for which the contract is set to last. For instance, if a $1000 contract is awarded for a duration of 10 months starting in January 2020, a nominal value of $100 is allocated for each month from January to October 2020.

As with the other data presented in this chapter, the expenditure values were deflated by the military expenditure deflator after aggregation[15]. Figure 3.20 compares the DoD obligations and the actual spending at a monthly frequency, while Figure 3.21 does the same at a yearly frequency.

Figure 3.20: DoD obligations vs DoD actual spending (monthly)

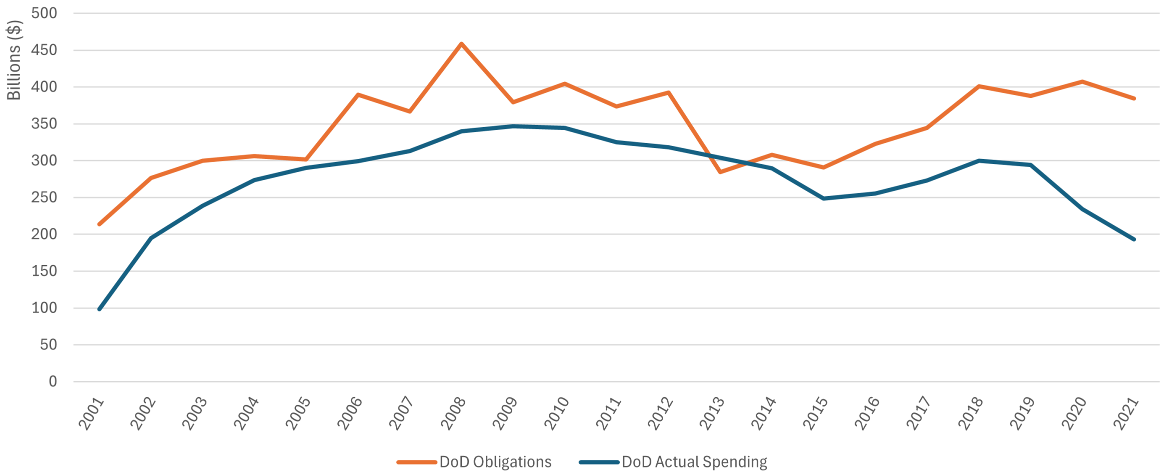


Notes: *The figure shows the evolution of DoD obligations and actual spending from 2001 to 2021 on a monthly basis. DoD obligations refer to the total contract values based on their signing date, while DoD actual spending allocates the nominal value of a contract over time according to its length. All values have been adjusted for inflation using the national defence government expenditure implicit price deflator and are presented in December 2021 dollars.*

that specific period.

[15]Note that the deflation leads to differing total values in DoD obligations and actual spending, particularly concerning high-value and long-term contracts. This occurs because, for the measurement of obligations, the entire procurement value is deflated by the index on that date, whereas for actual spending, each portion of the contract is deflated by the corresponding index value applicable at the time that spending is allocated.

Figure 3.21: DoD obligations vs DoD actual spending (annual)



Notes: *The figure shows the evolution of DoD obligations and actual spending from 2001 to 2021 on an annual basis. DoD obligations refer to the total contract values based on their signing date, while DoD actual spending allocates the nominal value of a contract over time according to its length. All values have been adjusted for inflation using the national defence government expenditure implicit price deflator and are presented in December 2021 dollars.*

Considerable volatility and seasonality can be observed when DoD obligations are presented monthly, as indicated in Figure 3.20. The figure also indicates that expenditure typically spikes in September, with March and December also showing higher values than average, supporting the information presented in Figure 3.4. When the aggregation of expenditure is spread over time using the contract length, the volatility and seasonality issue is significantly mitigated, as can be observed from the behaviour of the actual spending curve. However, it remains with some level of seasonality as some short-term procurements are awarded in months like January and September.

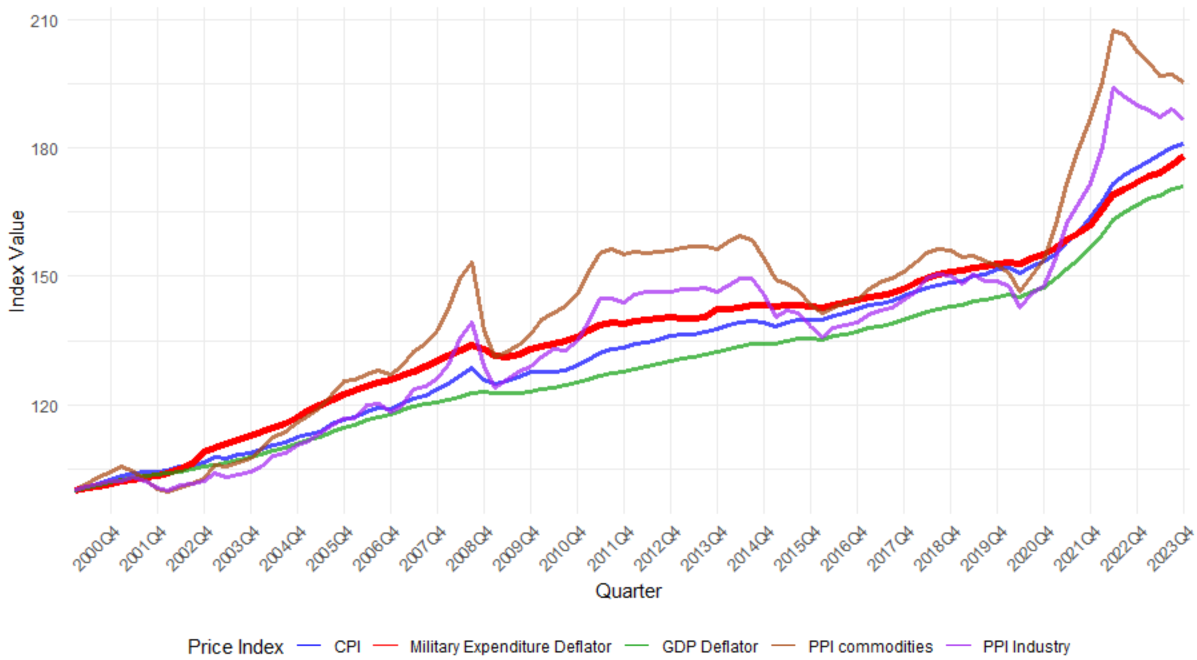
Notice from both figures that the initial years of the actual spending data exhibit lower values in comparison to the DoD obligations, as part of the expenditure from procurements signed in years preceding 2001 would likely be allocated to it and the other early years of the data range. As can be observed from Figure 3.21, the actual spending curve lies below the obligation curve in almost all instances, which can be attributed to long-length contracts, as a portion of their values is allocated to future years and some of the expenditure was allocated to years beyond 2021, which is after the last year of the data frame. From 2019, there is a noticeable increase in the gap between the two curves, indicating that a significant number of contracts with long durations and high values were awarded in 2020 and 2021.

As evidenced in the previous subsection, a significant portion of total procurement expenditure is derived from lengthy and high-value contracts. This implies that the actual spending will exhibit considerable persistence, as the nominal value of these extensive and costly procurements is evenly distributed over time. Furthermore, differentiating between obligations and actual spending may be more impactful when analysing data at a more granular level. Aggregated data typically exhibits less volatility, as demonstrated by comparing monthly and yearly expenditure patterns. For instance, if a location that rarely receives procurement contracts is awarded a sporadic long-term and costly contract, spreading the value of that contract over time will significantly affect the aggregated expenditure measurement.

3.3.8 Evolution of price indexes

Finalizing this section, this subsection contributes to the analysis of DoD expenditure by examining the evolution of the implicit price deflator for government expenditure in national defence alongside four other measures of inflation: the Consumer Price Index (CPI), the GDP deflator, the Producer Price Index (PPI) by commodity, and the PPI by industry. Figure 3.22 illustrates the evolution of these indices from 2001Q4 to 2021Q4, all re-based such that 2001Q1 = 100.

Figure 3.22: Evolution of price indexes



Notes: *The figure illustrates the evolution of five price indexes: the Consumer Price Index (CPI), the implicit price deflator for national defence government expenditure (military expenditure deflator), the GDP implicit price deflator, the Producer Price Index (PPI) by commodities, and the PPI by industry. These indexes are presented on a quarterly basis, covering the period from the first quarter of 2000 to the fourth quarter of 2021. All indexes have been re-based such that Q12001 equals 100.*

The military expenditure deflator closely follows the CPI and GDP deflator series. In the first seven quarters illustrated, the three indices are remarkably similar in value. From 2002Q4 to 2008Q4, a period characterised by an increase in military expenditure, the military expenditure deflator grew more rapidly than both the CPI and the GDP deflator, causing it to remain above these two indices for most of the period depicted in the figure, except at the end of 2021, when the CPI catches up to the military expenditure deflator. Despite these characteristics, the three indices display similar behaviour. However, this cannot be said for the PPI indices, which show considerably greater volatility, attributable to the fact that production costs might not be immediately passed on to consumers. The PPI tends to be more sensitive to input prices and supply chain dynamics.

3.4 Concluding remarks

This chapter has addressed the technical challenges involved in transforming highly granular procurement data into aggregate figures suitable for empirical analysis, offering a set of tools and R scripts designed to facilitate this task. It also presented an overview of a military expenditure dataset constructed from U.S. government procurement records, highlighting the dataset's key structural features. The use of military procurement data provides a valuable alternative to more traditional methods used in the empirical literature to identify expenditure shocks, while also offering the advantage of exceptional granularity across time, geography, sector, and other dimensions.

The DoD procurement expenditure in discussion is attractive for fiscal policy analysis because it is less influenced by policy responses to economic conditions than other forms of government spending. It is primarily driven by security-related objectives rather than macroeconomic performance. This characteristic lends it a higher degree of exogeneity. Additionally, procurement expenditure displays significantly greater volatility than both GDP and aggregate government expenditure.

Although the procurement dataset contains a large number of short-term, low-value contracts, the majority of total spending originates from a relatively small number of high-value, long-term contracts. This concentration implies that a substantial share of expenditure is directed to a limited group of firms and contributes to a high degree of persistence in the flow of procurement spending over time. When aggregating these expenditures, accounting for contract duration may be particularly important, especially as the level of disaggregation increases. The timing of payments and contract lengths can affect the interpretation of monthly or annual expenditure patterns.

Additionally, the composition of DoD procurement expenditure differs significantly from private sector spending, underlining the distinct nature of government-driven defence investments. The majority of the expenditure is allocated to the manufacturing sector, with the dominance of aerospace, missile, technology and vehicle manufacturer firms among the top contractors. Geographically, expenditure remains concentrated in a few key counties over time, largely due to the renewal of contracts with major firms. The geographical HHI presented in this chapter indicates that this concentration is increasing over time, especially in the manufacturing sector.

These features underscore the distinctive nature of military procurement expenditure and suggest that it may play a meaningful role in regional and sectoral economic development. The dataset introduced here can serve as a valuable foundation for further research into the broader economic effects of such spending, particularly in areas such as labour markets, innovation, and regional inequality. In the following chapter, I focus on one of these dimensions, analysing the effects of military procurement on labour market outcomes, both from a geographical and sectoral perspective.

Chapter 4

The regional and sectoral labour market effects of military procurement expenditure

4.1 Introduction

Government spending is a major component of aggregate expenditure and is commonly associated with stimulating economic activity, promoting growth, and improving public well-being. However, its effects on outcome variables of interest can vary significantly depending on features such as the type of spending, the method of financing, the targeted sectors, and the regional context in which it occurs. Analysing government expenditure through more disaggregated perspectives, such as regional and sectoral, can reveal dynamics that are not observable at the national level.

This chapter builds upon the data presented in the previous essay to examine the effects of DoD procurement expenditure on local labour markets. Unlike other forms of government spending, such as public employment or direct transfers, procurement spending, particularly in defence, injects capital into private-sector production, often independent of local business cycles. This relative exogeneity provides an opportunity to estimate the causal impacts of public spending on regional labour market outcomes, including employment, wages, and labour mobility.

A large body of research in macroeconomics and applied econometrics has focused on estimating the effects of government expenditure on aggregate outcomes, particularly GDP. This study contributes to that literature by shifting the focus to regional and sectoral dynamics, exploring the localised labour market effects of government spending, as well as the spatial and cross-sectoral spillovers that may arise from labour mobility, input-output linkages, or general equilibrium effects.

Estimating the impacts of government expenditure poses several challenges, the most notable being endogeneity. Common strategies to address this issue include narrative identification approaches, as in Ramey and Shapiro (1998), and structural vector autoregressions (SVARs) or sign-restriction methods, as in Blanchard and Perotti (2002). As discussed in the previous chapter, military expenditure, particularly procurement, proves very useful in this context because it is more strongly influenced by geopolitical considerations than by local or national business cycle fluctuations, thereby helping to mitigate endogeneity concerns.

Papers on fiscal policy and multipliers, such as Chodorow-Reich (2019) and Ramey (2011a), highlight essential insights relevant to this study. The impacts of government spending and the magnitude of estimated fiscal multipliers can vary significantly when examined at the regional level as opposed to the aggregate level, largely due to differing motivations. Two key reasons for this discrepancy are closely related to factors discussed in Chapter 2, specifically regarding the mapping of the regional wage Phillips curve to the national one.

The first reason for divergence between regional and aggregate effects is the presence of spillover effects, which are also examined in this chapter. While additional government expenditure in a county may lead to local job creation, its effects often extend beyond the targeted area. Factor mobility and relative price adjustments may allow firms in the receiving county to attract labour and inputs from neighbouring regions or sectors, potentially drawing economic activity away from them. At the same time, positive income effects can stimulate demand not only in nearby counties but also across sectors, increasing employment in industries that supply goods or services to the regions experiencing the initial boost in spending.

As a result, employment may rise not only in the county receiving the expenditure but also in adjacent counties and related sectors, though these spillover effects are not necessarily positive. The direction and magnitude of the spillovers depend on local labour market frictions, commuting patterns, and the structure of interregional and intersectoral linkages. For instance, Dupor and Guerrero (2017), using state-level data on defence procurement expenditure, find small but statistically significant spillover effects between U.S. states that are major trade partners. They argue that these inter-state spillovers account for a substantial share of the difference between regional and aggregate fiscal multipliers.

In a subsequent study, Dupor and McCrory (2018) use county-level data from the American Recovery and Reinvestment Act (ARRA) of 2009 and find further evidence of labour market spillovers from government spending. The analysis shows that expenditure in a given county leads to job creation and wage increases not only locally but also in nearby areas that belong to the same labour market, which is defined by commuting flows.

The second reason for divergence between regional and aggregate effects relates to endogenous policy responses, particularly from monetary authorities. At the national level, interest rates may respond to fiscal expansions, partially offsetting their effects on output and employment. However, such monetary policy responses are typically calibrated to aggregate economic conditions and are less relevant when evaluating regional impacts[1]. This makes regional analysis especially useful for isolating specific transmission channels through which government expenditure affects local labour markets.

Another critical factor that shapes the effects of fiscal policy and is fundamental to understanding the policy outcomes is the method of financing. From a regional standpoint, procurement spending can be considered externally financed, as it is funded by the federal government. This contrasts with the national perspective, where such spending affects the overall fiscal balance, and with other types of expenditure that may be financed locally. The source of funding matters: locally financed expenditure may generate weaker effects due to a Ricardian behaviour, as local residents may anticipate future tax increases and reduce their current consumption and investment in response.

In addition, many state and local governments in the U.S. are subject to balanced budget requirements, which, as Clemens and Miran (2012) note, often lead to pro-cyclical fiscal policy. During economic downturns, these constraints may force spending reductions precisely when stimulus is needed most. In contrast, military procurement is not subject to such rules, making it more stable and less responsive to local economic conditions. As a consequence, the effects of procurement spending on regional labour markets may be stronger, as it is less likely to be offset by contemporaneous fiscal contractions or anticipatory behavioural responses.

[1] Nakamura and Steinsson (2014) and Chodorow-Reich (2019) raise the discussion that aggregate multipliers are significantly larger in a lower zero bound situation, as interest rates become unresponsive to aggregate shocks. This scenario is often used in the literature as a benchmark for evaluating the strength of fiscal policy, since monetary policy is constrained. However, this comparison is not without limitations, as central banks may still influence macroeconomic conditions through tools such as quantitative easing and forward guidance, even when nominal interest rates are near zero.

While much of the existing literature on migration emphasises the connection between mobility and tax levels or the provision of public goods, drawing inspiration from Tiebout (1956) concept of "voting with their feet", this chapter presents a different perspective. Procurement spending does not impose fiscal burdens on local governments, nor does it necessitate compensatory tax increases; furthermore, it is not linked to the provision of local public goods. Consequently, this approach offers a distinct framework for understanding how public expenditure can influence labour market dynamics without triggering local budget constraints.

As previously mentioned, this chapter also investigates potential spillover effects from procurement expenditure across different sectors. A growing body of literature emphasises the importance of network effects in generating such spillovers, particularly through input-output linkages and supply chain interdependencies, which can amplify the effects of public spending across the broader economy. Acemoglu et al. (2016) show that both input-output and geographic networks play a key role in propagating localised shocks, leading to significant macroeconomic implications.

In the case of demand-side shocks, such as procurement expenditure, the authors find that the effects are predominantly upstream—increasing demand for suppliers of intermediate inputs—while downstream effects, such as rising costs for input users, tend to be much smaller or even negligible. This finding aligns with the evidence presented in this chapter: spending in the manufacturing and services sectors not only increases employment within those sectors but also generates positive spillover effects across all sectors of the economy.

Bouakez et al. (2023) also highlight the importance of sectoral heterogeneity and production networks, demonstrating that a New Keynesian model incorporating multiple sectors yields a fiscal multiplier that is approximately 75 percent larger than its one-sector counterpart. Other studies also use US procurement data disaggregated data to analyse the sectoral effects of spending. Using local projection estimations, Auerbach et al. (2020) find evidence of positive spillover effects from military procurement spending to the earnings and GDP of other industries in the same location. These effects arise from both upstream linkages and general equilibrium adjustments. In addition, they find no evidence of any crowding-out effects at nearby industries.

Similarly, Barattieri et al. (2023) examine the effects of procurement expenditure through the lens of production networks and sectoral heterogeneity. Also relying on local projection techniques, they document increases in employment, wages, and prices not only in the sectors directly receiving procurement funds, but also in input-supplying industries. Their findings further suggest that procurement in sectors with stickier prices and those located further downstream in the production chain tend to generate stronger aggregate effects.

In addition to the sectoral spillover effects discussed above, this chapter also presents evidence of significant population relocation as a response to procurement expenditure. The results indicate that in counties receiving at least $2.8 million in procurement spending within a year, a 1% increase in expenditure leads to an increase of approximately 0.42% in local employment and 0.36% in the local population. Furthermore, the analysis reveals geographical spillover effects, as increases in expenditure within a county's commuting zone are also associated with population growth in neighbouring counties.

When examining the impact on wages, the estimated effects are positive but more modest. A 1% increase in procurement expenditure is associated with an average 0.05% rise in local wages, while counties within the same labour market experience a spillover effect of around 0.02%. These findings are broadly consistent with Auerbach et al. (2020), who report positive spillover effects on earnings in cities located within a 50-mile radius of the procurement recipient.

This chapter departs from the traditional fiscal-policy focus on multipliers by estimating elasticities of the DoD procurement expenditure with respect to local labour market outcomes, such as employment, wages, and population. Methodologically, I partition the observations into expenditure groups using dummy variables, capturing non-linearities and separating intensive (high-spend) from extensive (low/zero-spend) margins, so that the full cross-sectional and time variation contributes to identification while very small values do not dilute the estimated effects.

At the sectoral level, this treatment is especially valuable given the prevalence of zeros and the skewness of contract values, yielding cleaner estimates of within- and cross-sector responses. Combined with a spatial framework and fixed effects, the approach provides a map from the effects of local procurement shocks to local labour market elasticities and their spillovers.

This chapter is organised as follows: Section 2 outlines the data sources and describes the empirical approach used for the analysis. Section 3 presents the results, including summary statistics of the data, and discusses and interprets the estimation findings. Section 4 provides a sectoral analysis of the impacts of military procurement expenditure on employment. Section 5 concludes the chapter.

4.2 Data and empirical approach

4.2.1 Data

The data measurements used in this chapter are listed below, along with their sources and brief descriptions. All is at the yearly frequency and represent the US counties, spanning from 2006 to 2021, with the exception of the Labour force, that ranges from 2007 to 2021.

Population The data was sourced by the United States Census Bureau.

Labour force Data originates from the Local Area Unemployment Statistics (LAUS), sourced by the Bureau of Labor Statistics (BLS). This measurement refers to the workers county of residence[2].

Employment Employment data is sourced from the Quarterly Census of Employment and Wages (QCEW). It denotes the total number of employed people in each county, regardless of the residence location.

[2] Counties in Connecticut were excluded from analyses that use the labour force as the outcome variable, because those data are available only under the state's redesigned county FIPS codes, making consistent matching with the rest of the dataset impossible.

Average weekly wages The data is also sourced from the QCEW. It denotes the average weekly wages paid by the companies at that location.

Military expenditure measurement The measurement used for the military expenditure data[3] is the "actual spending data", as described in Chapter 3. This measurement distributes procurement expenditure evenly over time based on its length. The expenditure data was adjusted using the national defence government expenditure implicit price deflator and is expressed in December 2021 dollars.

Commuting Zones and Labour Market Areas Commuting Zones (CMZ) and Labour Market Areas (LMA) are used to group counties into local labour markets, as in Chapter 2. The data is sourced from the Economic Research Service (ERS), which compiles the dataset using information from the Census. The referenced Commuting Zone data is derived from the 2020 Census, whereas the Labour Market Area data is based on the 1990 Census.

4.2.2 Empirical model

I initially examine the local impacts of procurement data by testing the following specifications for different labour market outcome variables:

$$y_{i,t} = \theta_s + \gamma_t + \beta_0 exp_{i,t} + \mu_{i,t} \tag{4.1}$$

$$y_{i,t} = \theta_s + \gamma_t + \beta_0 D^{high}_{i,t} + \beta_1 exp_{i,t} + \beta_2 D^{high}_{i,t} exp_{i,t} + \mu_{i,t} \tag{4.2}$$

The indices i, s, and t denote the county, state, and year, respectively. y represents an outcome variable, and exp denotes the expenditure. $D^{high}_{i,t}$ is a dummy variable that assumes the value one if the expenditure for county i is in the top 33% percentile during year t. θ and γ are fixed effects.

[3] Notice that in this chapter, I use the terms military/defence/DoD procurement expenditure and expenditure interchangeably.

Equation (4.1) captures the direct impact of expenditure on the outcome variables. Equation (4.2) is an augmented version of (4.1), which aims to capture heterogeneity in the impact of procurement expenditure across counties. The motivation for this specification is that defence procurement is highly unevenly distributed geographically. As discussed in the previous chapter and illustrated again in Table 4.4, many counties receive little or no procurement expenditure, while a much smaller group receives very large amounts. This feature is visible in the summary statistics, where the mean and upper percentiles of procurement expenditure are substantially larger than the median. In this context, imposing a single constant elasticity across all counties may be too restrictive, since an increase in procurement expenditure in a low-expenditure county may not have the same labour-market implications as an increase in a a high-expenditure county. The latter might be more likely to have the infrastructure needed to translate additional procurement expenditure into local employment, labour-force, and population responses, also generating stronger local multiplier effects.

Notice that the purpose of equation (4.2) is not to divide the sample into unrelated groups, but to allow the relationship between procurement expenditure and labour-market outcomes to differ across the expenditure distribution. The high-expenditure dummy identifies county-year observations in the upper third of the procurement distribution[4]. The threshold should therefore be interpreted as a parsimonious way to capture non-linearity in a highly skewed distribution, rather than as a structural cut-off between two distinct types of counties.

To capture geographical spillover effects of procurement expenditure, I expand on the previous two equations to include the local impacts resulting from expenditures in neighbouring areas. Since the focus is on local labour markets, I use CMZs to determine these markets. The equations tested are as follows:

$$y_{i,t} = \theta_s + \gamma_t + \beta_0 exp_{i,t} + \beta_1 exp_{i,t}^{CMZ} + \mu_{i,t} \tag{4.3}$$

[4]Notice that the threshold of 33% is arbitrary, but results are qualitatively similar when alternative cut-offs are used.

$$y_{i,t} = \theta_s + \gamma_t + \beta_0 D_{i,t}^{high} + \beta_1 D_{i,t}^{high,CMZ} + \beta_2 exp_{i,t} + \beta_3 D_{i,t}^{high} exp_{i,t} + \beta_4 exp_{i,t}^{CMZ} + \beta_5 D_{i,t}^{high,CMZ} exp_{i,t}^{CMZ} + \mu_{i,t} \tag{4.4}$$

The indices are the same as in the previous equations, with the addition of CMZ, which denotes the cluster defining the local labour markets[5]. The expenditure and outcome variables in a local labour market is composed by all the counties in that CMZ, excluding the county of interest itself. Similarly, $D_{i,t}^{high,CMZ}$ denotes if the expenditure in that area is in the top 33% percentile during year t. Equation (4.3) captures spillover effects, while (4.4) does the same but allows for distinct impacts for different expenditure groups as in (4.2).

In all specifications, year fixed effects are included to control for common shocks affecting all counties in a given year, such as aggregate labour-market conditions and national macroeconomic fluctuations. State fixed effects are included to absorb time-invariant differences across states, while preserving cross-county variation in procurement exposure within each state. By contrast, county fixed effects remove all time-invariant county characteristics and identify the coefficient exclusively from changes within the same county over time. The state-FE specification can be interpreted as capturing the relationship between procurement exposure and local labour-market outcomes across counties within the same state, rather than relying exclusively on short-run deviations from a county's own long-run procurement level. As a robustness exercise, the regressions are also estimated with county fixed effects, and their results are presented in the Appendix E.

All variables underwent an inverse hyperbolic sine transformation (arcsinh). The standard approach typically involves taking the logarithm of the data; however, because military expenditure includes many true zeros, using logarithms would require disregarding all observations that are equal to zero. An alternative method is to apply $log(a + x)$, assuming a value of one or an even smaller non-negative value for a.

The arcsinh transformation is becoming a popular technique as it approximates the

[5] An alternative set of results that uses LMA instead of CMZ to the determine local labour markets is presented in the Appendix D

logarithmic function for moderately large values, which is representative of most non-zero observations in many datasets, as in the case of procurement expenditure. Unlike logarithmic transformations, the arcsinh function is defined across the entire real line and does not require an arbitrary shift, as observed with $log(a + x)$.

Furthermore, as illustrated by Bellemare and Wichman (2020), using an arcsinh-arcsinh specification, where both the independent and dependent variables are transformed with the inverse hyperbolic sine, leads to the following elasticity of the outcome variable y with respect to the independent variable x:

$$\hat{\xi}_{yx} = \hat{\beta}\frac{\sqrt{y^2+1}}{y}\frac{x}{\sqrt{x^2+1}},$$

Where $\lim_{y\to\infty} \frac{\sqrt{y^2+1}}{y} = 1$, and $\lim_{x\to\infty} \frac{x}{\sqrt{x^2+1}} = 1$. So, for sufficiently large values of y and x, the regression coefficients can be interpreted as elasticities, akin to a log-log model. For values larger than 10, the bias is negligible. If the independent variable is small, the elasticity shows a bias toward zero, whereas if the outcome variable is small, there is a positive bias in the magnitude of the elasticity.

4.3 Results

4.3.1 Employment-related measures

This section begins by analysing the effects of the defence procurement expenditure on variables related to employment and the movement of people across counties. Equations (4.1) and (4.2) are estimated for the following outcome variables: Employment, Labor Force, and Population. Increased expenditure is expected to create new jobs and thereby impact these variables, however, it is important to highlight what changes in each of them represent.

The QCEW employment data indicates the number of individuals employed by local businesses, regardless of their residential location. When expenditure leads to the creation of new jobs, whether directly or indirectly, the resulting increase in employment can be split into different categories of new workers. Some of these individuals are drawn from the local surplus workforce, consisting of residents in the county who were previously unemployed or inactive. Additionally, some new workers come from other areas, either by relocating to the county or commuting to work.

The LAUS Labour Force data measures a distinct aspect. It is based on the worker's place of residence. In this context, new entrants to the local labour force are defined as residents of the county. These individuals may include local residents who were previously inactive but have opted to enter the labour force due to emerging job opportunities, as well as workers migrating from other counties.

In essence, population growth can be understood more straightforwardly. An increase in population resulting from shifts in expenditure can be attributed to migration to that location. However, this growth includes not only workers but also individuals who are not part of the labour force, such as family members who accompany those relocating for employment. Furthermore, it's important to consider that expenditures can trigger multiplier effects through income channels, meaning that changes in population may also indicate the arrival of new workers migrating to engage in the informal sector. Although the number of these individuals may be relatively small, they are not captured in the other two statistics.

A limitation of the empirical approach is that none of the outcome variables mentioned above accounts for another potential adjustment margin to the expenditure shock: an increase in the number of hours worked by currently employed individuals. Capturing this effect would require data on hours worked, which are unavailable in the current dataset.

When analyzing the results, it is crucial to consider the absolute size of each variable. The estimated coefficients from equations (4.1)–(4.4) represent elasticities, meaning their magnitudes are influenced by the base values of the outcome variables. For example, a 1% increase in population is not equivalent to a 1% increase in employment, as the former represents a significantly larger absolute change. To contextualize the estimation results, Tables 4.1, 4.2, and 4.3 present summary statistics for the outcome variables. These tables report the mean, standard deviation, minimum, maximum, and the 25th, 50th (median), and 75th percentiles. The statistics are shown for two years, 2007 and 2021, which correspond to the initial and final years of the dataset. Additionally, the statistics are categorized by low (bottom 66%) and high (top 33%) expenditure groups, in accordance with the classifications defined by the dummy variables in equations (4.2) and (4.4).

Table 4.1: Summary statistics: employment

Year	**Group**	**Mean**	**Std. Dev.**	**Min**	**25%**	**Median**	**75%**	**Max**
2007	*Total*	36130	130193	35	2414	6772	20321	3619392
	Low expenditure	6630	8314	35	1631	3741	8589	73507
	High expenditure	92785	210988	449	11232	31503	83581	3619392
2021	*Total*	38453	139394	40	2267	6491	20351	3702461
	Low expenditure	6934	9516	40	1524	3731	8714	116987
	High expenditure	98986	225902	397	10201	31644	90831	3702461

Notes: The table presents summary statistics for the distribution of county-level employment data in 2007 and 2021. The second column indicates the group of counties: "Total" includes all counties; "Low expenditure" refers to counties in the bottom 66% of military procurement expenditure; and "High expenditure" refers to counties in the top 33%. The dataset includes a total of 3079 counties.

Table 4.2: Summary statistics: labour force

Year	**Group**	**Mean**	**Std. Dev.**	**Min**	**25%**	**Median**	**75%**	**Max**
2007	*Total*	48281	155581	41	5202	12150	31713	4911024
	Low expenditure	11873	13083	41	3805	7799	15470	134369
	High expenditure	119513	252185	1018	19876	48332	117200	4911024
2021	*Total*	51232	168208	72	4653	11326	30827	4999687
	Low expenditure	11654	13856	72	3395	7242	14676	179711
	High expenditure	128664	272502	508	17556	47612	129597	4999687

Notes: The table presents summary statistics for the distribution of county-level labour force data in 2007 and 2021. The second column indicates the group of counties: "Total" includes all counties; "Low expenditure" refers to counties in the bottom 66% of military procurement expenditure; and "High expenditure" refers to counties in the top 33%. The dataset includes a total of 3079 counties.

Table 4.3: Summary statistics: population

Year	**Group**	**Mean**	**Std. Dev.**	**Min**	**25%**	**Median**	**75%**	**Max**
2007	*Total*	97411	311079	268	11412	26006	66598	972796
	Low expenditure	24538	26467	268	8174	16390	31724	301989
	High expenditure	237365	501634	2498	40826	97704	233902	9727968
2021	*Total*	107326	335781	256	11323	26394	69647	9809462
	Low expenditure	26161	29804	256	8189	16889	33204	422636
	High expenditure	263206	539258	842	39515	105686	265537	9809462

Notes: The table presents summary statistics for the distribution of county-level population data in 2007 and 2021. The second column indicates the group of counties: "Total" includes all counties; "Low expenditure" refers to counties in the bottom 66% of military procurement expenditure; and "High expenditure" refers to counties in the top 33%. The dataset includes a total of 3079 counties.

Upon comparing the mean values presented in the tables above, it is evident that the population figures significantly exceed both employment and the labour force, with the latter generally accounting for approximately 45–50% of the total population. Additionally, it is not surprising that all three variables show elevated values within the high expenditure group, highlighting that counties that receive greater amounts of expenditure typically have larger populations, labour forces, and employment levels on average.

Table 4.4 further details the regression data, including summary statistics for expenditure in a similar format. The statistics are also displayed for 2007 and 2021, with additional figures for the average of the expenditure across all years of the dataset included as well.

Table 4.4: Summary statistics: expenditure

Year	Mean	Std. Dev.	25%	33%	Median	66%	75%	Max
2007	72247022	419754348	0	21109	271487	**2045013**	5294642	10031323804
2021	61143720	385908263	0	422	230653	**1754857**	4956685	9723626334
Average	85033997	541931723	38509	119584	682275	**2881678**	6584861	13315558133

Notes: The table presents summary statistics for the distribution of county-level military procurement expenditure data in 2007 and 2021. The dataset includes a total of 3079 counties.

The table presented above includes the 66th percentile value, which highlights the expenditure level used to categorize the counties into different expenditure groups. It is noteworthy that both the 66th percentile value and the average are significantly higher than the median, suggesting a substantial concentration of expenditures among a subset of counties, while others receive considerably lower amounts.

The table below display the results from the estimation of equations (4.1) and (4.2) for each outcome variable:

Table 4.5: Estimation results
Employment, labour force, and population

Outcome variable	Employment		Labour force		Population	
Specification	**(4.1)**	**(4.2)**	**(4.1)**	**(4.2)**	**(4.1)**	**(4.2)**
Expenditure	0.127***	0.074***	0.112***	0.066***	0.110***	0.066***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)
D^{high} * Expenditure		0.345***		0.317***		0.297***
		(0.016)		(0.015)		(0.014)
D^{high}		-4.989***		-4.664***		-4.326***
		(0.265)		(0.250)		(0.232)
Counties	3087	3087	3122	3122	3130	3130
Years	16	16	15	15	16	16
Observations	49392	49392	46830	46830	50080	50080

Notes: The table presents the estimation results for equations (4.1) and (4.2), with employment, labour force, and population as the outcome variables. The first row indicates the outcome variable, while the second row the estimated equation. All variables in the regression have been transformed using an inverse hyperbolic sine transformation. Each regression includes a varying number of observations based on the outcome variable, as noted in the last row. All regressions account for both time and state fixed effects, and standard errors are clustered at the county level. The symbols "***", "**", and "*" indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

The findings presented in Table 4.5 illustrate the effects of increasing local defence procurement spending, which are in line with the expected results: new jobs are created, which in turn leads to a rise in employment, the size of the labour force, and population figures. Analysing the results of equation (4.2), we observe significantly higher coefficients for counties with expenditures in the top percentile across all three outcome variables. Specifically, these magnitudes are 0.419 for employment, 0.383 for the labour force, and 0.363 for population. Therefore, a 1% increase in expenditure results in increases of 0.42%, 0.38%, and 0.36% in these respective variables.

It is important to note that the impact on employment is more pronounced compared to the other two variables. The magnitudes of these effects may vary for several reasons. First, as previously mentioned, each variable captures a unique aspect of the labour market. Additionally, the absolute size of employment is typically smaller than that of the labour force[6], which, in turn, is smaller than the total population. This relationship is illustrated in Tables 4.1, 4.2, and 4.3. Consequently, the estimated coefficients will inherently differ in magnitude, as they represent percentage changes relative to various

[6]In some rare instances, the level of employment may exceed that of the labour force, as employment refers to the workers hired by local firms, while the labour force pertains to the residents of the county.

base values.

The findings demonstrate significant effects across all outcome variables. However, considering the substantial base value for the population, it appears that the majority of new job positions generated by increased expenditure are likely to be filled by workers relocating from other counties. This outcome highlights a considerable degree of labour mobility within the country.

These findings should be interpreted alongside the alternative specifications with county fixed effects, reported in Appendix E. The county-FE estimates are considerably smaller than the baseline estimates. For employment, labour force, and population, the coefficient on expenditure is effectively zero across specifications. The interaction between high-expenditure counties and procurement remains statistically significant for employment and population in some specifications, but the magnitudes are very small compared with the baseline estimates. These results indicate that the large baseline effects in Table 4.5 are not primarily driven by short-run changes in procurement within the same county over time.

The difference between the state-FE and county-FE estimates reflects the different sources of identifying variation used by the two specifications. The state-FE specification compares counties within the same state and therefore preserves persistent differences in procurement exposure across counties. By contrast, the county-FE specification compares each county only to itself over time, removing all time-invariant county characteristics and identifying the coefficient only from deviations from the county's own historical procurement level. The near-zero county-FE estimates therefore suggest that the baseline relationship is mainly associated with persistent differences in procurement exposure across counties, rather than with annual within-county fluctuations in procurement military expenditure.

The positive coefficients reported in Table 4.5 should be read as evidence that counties with persistently higher defence-procurement exposure tend to have higher employment, labour force, and population levels relative to other counties in the same state. They should not be interpreted as evidence that annual increases in procurement generate similarly large short-run changes within the same county.

The following table analyses local spillover effects and presents the estimation results for equations (4.3) and (4.4), using the same outcome variables as shown in Table 4.5:

Table 4.6: Estimation results (spillover effects)
Employment, labour force, and population

Outcome Variable	**Employment**		**Labour Force**		**Population**	
Specification	**(4.3)**	**(4.4)**	**(4.3)**	**(4.4)**	**(4.3)**	**(4.4)**
Expenditure	0.124*** (0.003)	0.071*** (0.003)	0.108*** (0.003)	0.063*** (0.002)	0.103*** (0.002)	0.060*** (0.002)
D^{high} * Expenditure		0.339*** (0.016)		0.305*** (0.015)		0.288*** (0.014)
Expenditure CMZ	0.029*** (0.004)	0.017*** (0.004)	0.037*** (0.003)	0.021*** (0.003)	0.036*** (0.003)	0.022*** (0.003)
$D^{high,CMZ}$ * Expenditure CMZ		0.064*** (0.023)		0.081*** (0.023)		0.062*** (0.022)
D^{high}		-4.899*** (0.265)		-4.497*** (0.246)		-4.189*** (0.223)
$D^{high,CMZ}$		-1.245** (0.484)		-1.501*** (0.481)		-1.120** (0.447)
Counties	3080		3122		3123	
Years	16		15		16	
Observations	49280		46830		49968	

Notes: The table displays the estimation results for equations (4.3) and (4.4), focusing on employment, the labour force, and population as the outcome variables. The first row of the table specifies the outcome variable, while the second row presents the estimated equation. All variables in the regression have been transformed using an inverse hyperbolic sine transformation. Commuting zones (CMZ), as per the 2020 census data, were used to define local labour markets. The variables denoted by CMZ represent the values for the CMZ the county belongs to, excluding itself. Each regression includes a varying number of observations based on the outcome variable, as noted in the last row. All regressions account for both time and state fixed effects, and standard errors are clustered at the county level.
The symbols "***", "***", and "*" indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

The estimated local expenditure effects are broadly consistent with the results from equations (4.1) and (4.2), as shown by the coefficients on $Expenditure$ and $D^{high} * Expenditure$. The results also provide evidence of spillovers within local labour markets. Across the three outcome variables and both specifications, procurement expenditure in other counties within the same commuting zone is positively associated with local employment, labour force, and population.

For counties located near high-expenditure areas, the estimated spillover effects are larger. A 1% increase in procurement expenditure in neighbouring counties within the same CMZ is associated with a 0.08% increase in local employment, a 0.10% increase in the labour force, and a 0.09% increase in population. These patterns are consistent with two spillover channels operating simultaneously. First, higher procurement expenditure may increase labour demand in the directly affected counties, attracting workers from nearby areas and potentially weakening labour-market outcomes in neighbouring counties. Second, procurement may stimulate the wider regional economy through commuting links, supplier relationships, household spending, and local demand spillovers. In this case, neighbouring counties may also benefit, especially if workers choose to live in adjacent areas while working in the high-expenditure county, or if firms in surrounding counties participate indirectly in procurement-related activity.

When comparing these results with the LMA-based estimates reported in Table D1 in Appendix D, some differences emerge. In equation (4.3), the CMZ specification produces positive spillover coefficients across all three outcome variables. By contrast, the LMA-based spillover estimates are generally smaller and are not statistically significant for employment. In equation (4.4), the LMA specification produces small negative spillover coefficients for lower-expenditure areas, although these effects are offset for counties neighbouring high-expenditure areas. Overall, the estimated spillovers are weaker when LMAs are used instead of CMZs.

This difference is likely related to the way the two geographical units are defined. CMZs are designed to capture commuting-based local labour-market linkages and tend to represent tighter economic areas. LMAs, by contrast, are larger units constructed around a minimum population threshold and are therefore more geographically expansive. As a result, CMZs may better capture the local labour-market interactions through which procurement spillovers operate. The stronger CMZ results suggest that the relevant spillover channels are relatively local and are more likely to occur among counties with stronger commuting and economic interdependence.

The results therefore suggest that spillovers are heterogeneous and potentially nonlinear. At lower levels of neighbouring expenditure, spillover effects are relatively small and may be partly offset by worker reallocation toward high-expenditure counties. However, when neighbouring procurement exposure is high, wider regional demand effects appear strong enough to generate positive net spillovers for surrounding counties. This interpretation is consistent with the stronger spillover estimates obtained using CMZs, where counties are more closely connected through local labour-market relationships.

Employment per capita To enhance the analysis, this section presents the estimation results for the same regressions as previously conducted, but focuses on the impacts of the variables on per capita levels. In this case, per capita employment serves as the outcome variable, and expenditure is also expressed in per capita terms. Additionally, the dummy variable indicating the expenditure groups is constructed based on per capita expenditure.

This test has two primary objectives. First, it accounts for the size of the county population. Second, it provides insights into the relative effects on employment and population. Those were not clear from the previous results, as the estimated elasticities refer to different base values. Table 4.7 showcases the estimation outcomes:

Table 4.7: Estimation results
Outcome variable : Employment per capita

Specification	**(4.1)**	**(4.2)**	**(4.3)**	**(4.4)**
PC Expenditure	0.011*** (0.001)	0.010*** (0.001)	0.011*** (0.001)	0.011*** (0.001)
D^{high} * PC Expenditure		0.000 (0.003)		0.001 (0.002)
PC Expenditure CMZ		0.010 (0.014)	-0.003*** (0.001)	-0.002* (0.001)
$D^{high,CMZ}$ * PC Expenditure CMZ				0.002 (0.004)
D^{high}				0.013 (0.013)
$D^{high,CMZ}$				-0.027 (0.026)
Counties	3087			
Years	16			
Observations	49392			

Notes: The table displays the estimation results for equations (4.1) to (4.4), using employment per capita as the outcome variable. The estimated equation is indicated in the first row of the table. Notice that equations (4.3) and (4.4) were estimated using Commuting Zones (CMZ), as per the 2020 census data, to determine the local labour market. The variables denoted by CMZ represent the values for the CMZ the county belongs to, excluding itself. All variables in the regression have been transformed using an inverse hyperbolic sine transformation. Each regression includes a varying number of observations based on the outcome variable, as noted in the last row. All regressions account for both time and state fixed effects, and standard errors are clustered at the county level.
The symbols "***", "**", and "*" indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

A notable feature of the results, as observed in the table, is that when using per capita values, there is no significant difference in the estimated effects between counties with low and high expenditure levels [7]. It is important to note that the reported coefficients in this specification cannot be directly interpreted as elasticities. As previously discussed, interpreting coefficients as elasticities requires the absolute values of both the dependent and independent variables to be sufficiently large—typically on the order of 10 or higher. This condition is not satisfied for employment per capita, which ranges between 0 and 1.

As outlined in the previous section, a correction is necessary to interpret these cases as elasticities. Since the average employment per capita in the dataset is approximately

[7] Notice also from Table E3, that no statistically significant effects are found when the specification controls for county FE.

0.356, all estimated coefficients must be multiplied by a scaling factor of 2.98 to derive the elasticities. This factor is derived from $\frac{\sqrt{y^2+1}}{y}$, where y represents the average per capita employment.

Even after the correction, the impact on per capita employment remains relatively modest, which is understandable given that it pertains to a per capita value. The effect of local expenditure hovers around 0.03, implying that for every 1% increase in per capita expenditure, local per capita employment rises by 0.03%. This suggests that, on average, employment is grows slightly more than the population in the occurrence of an expenditure raise.

However, the results present a contrasting picture regarding spatial spillover effects. There is a small, yet negative, spillover effect on per capita employment when expenditure in neighbouring labour markets increases. This effect is more pronounced when LMAs are used as clusters, as can be observed in Table D2. These findings indicate that in such regions, the population expands at a rate that outpaces job growth. One possible explanation for this trend could be that workers are relocating to a particular county to commute and work in the neighbouring county that received the increased expenditure.

4.3.2 Effects on the level of wages

I now turn to the effects of expenditure on wage levels, repeating the previous analysis but using nominal wages as the outcome variable. If the increase in expenditure boosts local demand for workers, it may pressure local wages upwards, also possibly resulting in spatial spillover effects as the same firms will also compete for workers in the neighbouring regions. As discussed in the data section, the measurement of wages available at the county-level and hence used for the analysis is average weekly wages. Table 4.8 presents the summary statistics for this variable.

Table 4.8: Summary statistics: Average weekly wages

Year	**Group**	**Mean**	**Std. Dev.**	**Min**	**25%**	**Median**	**75%**	**Max**
2007	*Total*	599	152	300	503	569	654	2156
	Low expenditure	557	123	300	484	537	603	1614
	High expenditure	679	170	378	573	640	746	2156
2021	*Total*	903	230	421	773	858	973	3761
	Low expenditure	849	173	421	749	822	914	2259
	High expenditure	1006	285	454	843	951	1089	3761

Notes: The table presents summary statistics for the distribution of county-level average weekly wages data in 2007 and 2021. The second column indicates the group of counties: "Total" includes all counties; "Low expenditure" refers to counties in the bottom 66% of military procurement expenditure; and "High expenditure" refers to counties in the top 33%. The dataset includes a total of 3079 counties.

Notice from the table that hourly wages are also higher, on average, in the counties belonging to the high expenditure group, also displaying higher volatility in that group. Table 4.9 presents the estimation results for equations (4.1) to (4.4) using wages as the dependent variable.

Table 4.9: Estimation results
Outcome variable: average weekly wages

Specification	**(4.1)**	**(4.2)**	**(4.3)**	**(4.4)**
Expenditure	0.008*** (0.000)	0.003*** (0.000)	0.008*** (0.000)	0.002*** (0.000)
D^{high} * Expenditure		0.045*** (0.003)		0.043*** (0.003)
Expenditure CMZ			0.003*** (0.001)	0.001*** (0.001)
$D^{high,CMZ}$ * Expenditure CMZ				0.017*** (0.004)
D^{high}		-0.690*** (0.046)		-0.660*** (0.045)
$D^{high,CMZ}$				-0.315*** (0.091)
Counties	3087		3080	
Years	16		16	
Observations	49392		49280	

Notes: The table displays the estimation results for equations (4.1) to (4.4), using average wages as the outcome variable. The estimated equation is indicated in the first row of the table. Notice that equations (4.3) and (4.4) were estimated using Commuting Zones (CMZ), as per the 2020 census data, to determine the local labour market. The variables denoted by CMZ represent the values for the CMZ the county belongs to, excluding itself. All variables in the regression have been transformed using an inverse hyperbolic sine transformation. Each regression includes a varying number of observations based on the outcome variable, as noted in the last row. All regressions account for both time and state fixed effects, and standard errors are clustered at the county level.
The symbols "***", "**", and "*" indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

The table above reveal a difference in results when focusing on counties in the high-expenditure percentile. The local effects are consistent across specifications, with results from equations (4.1) and (4.3) aligning closely, as do those from equations (4.2) and (4.4). In low-expenditure counties, a 1% increase in procurement expenditure raises average wages by approximately 0.01%, while in high-expenditure counties, the effect is around 0.05%. One possible explanation for the relatively small magnitude of these wage effects lies in the findings from the previous outcome variables. While the expenditure shock increases labour demand, the resulting upward pressure on wages is partially mitigated by an inflow of workers into these counties, helping to meet the rising demand for labour.

Although these percentage changes may appear modest, they can translate into substantial wage increases in counties experiencing significant fluctuations in procurement expenditure. For instance, among the top 33% of counties by expenditure, the average annual growth rate over the sample period is approximately 12.9%. This corresponds to an average increase in local wages of around 0.65%. Moreover, in the same group, expenditure swings exceeding 50% were observed on 4,270 occasions across the dataset. Such large shocks would be associated with local wage increases of at least 2.5%.

The results for equations (4.3) and (4.4) also provide evidence of small spatial spillover effects on wages. For low-expenditure counties, the spillover effect is insignificant when using CMZs and negative but negligible in magnitude under LMAs. In high-expenditure counties, positive wage spillovers of approximately 0.02% per 1% increase in expenditure are observed. Positive spillover effects are expected on wages, as the expenditure shock increases regional labour market tightness as firms compete for workers in all the neighbouring areas.

It is important to note that the wages in the regressions are measured in nominal terms. To some extent, the estimated equations capture effects on real wages, as the inclusion of time fixed effects accounts for price variations common to all counties, effectively controlling for national inflation. However, they do not account for price differences across counties. A county that receives an expenditure shock is likely to experience relatively higher local prices, meaning the estimated effects may overstate the true impacts on real wages. Fully assessing these effects would require county-level price data to capture local inflation, but such data are unavailable.

Notice also that the county FE estimates provided in the Appendix E display a different pattern from those observed for employment, labour force, and population. While the inclusion of county fixed effects reduces the estimated effects on the quantity variables to values close to zero, the wage regressions still show statistically significant coefficients for local effects. However, the sign and interpretation of the high-expenditure interaction change. In the county fixed-effect specification, procurement expenditure is positively associated with average weekly wages, but the interaction between high-expenditure status and expenditure is negative. This suggests that, once permanent county characteristics are removed, short-run increases in procurement are associated with higher wages mainly outside the highest-expenditure county-years, while the marginal wage effect is weaker in counties already receiving very high levels of procurement. In lower-expenditure counties, additional short-run procurement expenditure may represent a more unusual demand shock and therefore place higher upward pressure on local wages.

Asymmetric effects on wages. Building upon a pertinent discussion in the literature concerning downward wage rigidities, I extend the analysis of the impact of expenditure on wage levels to explore potential asymmetries in the effects of both positive and negative expenditure shocks. To achieve this, I augment equations (4.1) and (4.2) by introducing a dummy variable that takes the value of one if a county's expenditure in a given year has increased relative to the prior period, and zero otherwise. This specification enables the model to estimate distinct coefficients for periods of increasing and declining expenditure, thereby capturing any asymmetries in wage responses. The estimated equations are as follows:

$$y_{i,t} = \theta_s + \gamma_t + \beta_0 D_{i,t}^{pos} + \beta_1 exp_{i,t} + \beta_2 D_{i,t}^{pos} exp_{i,t} + \mu_{i,t} \tag{4.5}$$

$$\begin{aligned} y_{i,t} = \theta_s + \gamma_t + \beta_0 D_{i,t}^{pos} + \beta_1 D_{i,t}^{high} + \beta_2 D_{i,t}^{pos} D_{i,t}^{high} + \beta_3 exp_{i,t} + \\ \beta_4 D_{i,t}^{pos} exp_{i,t} + \beta_5 D_{i,t}^{high} exp_{i,t} + \beta_6 D_{i,t}^{pos} D_{i,t}^{high} exp_{i,t} + \mu_{i,t} \end{aligned} \tag{4.6}$$

The dummy $D_{i,t}^{pos}$ assumes value 1 if the growth rate of the expenditure of the county i in year t was non-negative and zero otherwise. All the remaining variables are the same as in specifications (4.1) and (4.2)[8].

[8] For simplicity, spillover effects are not included in this analysis.

Table 4.10: Estimation results
Asymmetric effects on wages

Specification	**(4.5)**	**(4.6)**
Expenditure	0.008***	0.003***
	(0.002)	(0.000)
D^{pos} * Expenditure	0.012**	-0.001
	(0.001)	(0.001)
D^{high} * Expenditure		0.046***
		(0.003)
D^{pos} * D^{high} * Expenditure		0.001
		(0.002)
D^{pos}	-0.200***	-0.004
	(0.017)	(0.017)
D^{high}		-0.711***
		(0.052)
D^{pos} * D^{high}		-0.023
		(0.032)
Counties	3087	
Years	15	
Observations	46305	

Notes: The table displays the estimation results for equations 5 and 6, using average wages as the outcome variable. The estimated equation is indicated in the first row of the table. All variables in the regression have been transformed using an inverse hyperbolic sine transformation. All regressions account for both time and state fixed effects, and standard errors are clustered at the county level. The symbols "***", "**", and "*" indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

The findings from specification (4.5) reveal a small asymmetry in the results: a decrease in expenditure does not yield a wage deflation of the same magnitude as the inflationary effect of an increase in expenditure. Specifically, a 1% reduction in expenditure results in an average wage increase of approximately 0.01%, whereas a 1% increase in expenditure leads to a wage increase nearly double that amount, around 0.02%.

However, this evidence of asymmetry in the effects disappears upon the introduction of the dummy variable that differentiates counties based on expenditure groups. As seen in the estimated results of equation (4.6), the coefficients that interact with D_{it}^{pos} are not statistically significant, and the implied coefficients remain consistent with those reported in Table 4.9.

The results indicate that the key factor in determining the impact on wages is whether the expenditure size is sufficiently high. Additionally, the correlation between the two dummy variables employed is 0.246, signifying a weak positive association. This suggests that counties classified in the high expenditure group are somewhat more likely to see an increase in expenditure compared to their counterparts. Consequently, the variable D^{high} captures some of the effects related to the asymmetric influences on wage levels.

Thus, this analysis does not provide compelling evidence of asymmetric effects on wages linked to procurement expenditure. It is also important to consider that since average weekly wages are used as the measurement, any wage increases could potentially stem from a rise in the average number of hours worked. If the shock is temporary, wages may subsequently decline at a similar rate as workers reduce their hours.

4.4 Sectoral analysis

This section shifts the analysis to the sectoral dimension, complementing the previously examined effects of procurement expenditure on regional labour market outcomes and their potential spatial spillovers. The purpose of this section is to identify and analyse sector-specific heterogeneity in responses to expenditure shocks. The empirical examination is restricted to employment, as population data cannot be segmented at the sectoral level, and the labour force data used in this chapter is similarly not available at that level of detail. Additionally, the impact on wages is not addressed, given the small magnitudes observed in the previous section, likely due to a high degree of labour mobility.

I examine the data aggregated into four primary sectors: Manufacturing, Services, Construction and Natural Resources. Counties with missing data in any sector[9] were excluded from the analysis to ensure that the regressions for each sector feature the same number of observations and encompass the same counties, thus allowing for comparability

[9] In some cases, employment data for certain sectors/counties are not disclosed by the QCEW due to confidentiality issues.

of the results. Following this filtering process, the dataset contains a total of 2,903 counties.

The allocation of employment data into sectors is straightforward, as the QCEW already provides employment figures disaggregated into the four primary sectors used in this analysis. In fact, the QCEW offers a more aggregated division into two categories: service-producing and goods-producing industries. The former corresponds to the services sector, which is further broken down into numerous subsectors (not used in this analysis). The latter includes three subcategories, which are natural resources and mining, construction, and manufacturing, which together make up the other three sectors analysed in this study.

Procurement expenditure was classified into the four primary sectors using their corresponding NAICS 2-digit codes. The NAICS system defines 21 distinct groups at the 2-digit level, which were allocated to the four primary sectors accordingly. Table F1 in Appendix F provides a detailed mapping of each 2-digit NAICS category to its assigned primary sector. This sectoral classification of the procurement expenditure is consistent with the classification used for employment, as the QCEW also relies on NAICS codes for its categorisation.

The following tables present descriptive statistics to illustrate the sectoral distributions of the employment and expenditure across counties. Table 4.11 displays the distribution of the sectoral employment for the year of 2021.

Table 4.11: Summary statistics: distribution of sectoral employment (2021)

Sector	**Share of national employment**	**Mean**	**Std. Dev.**	**Min**	**25%**	**Median**	**75%**	**Max**
Manufacturing	**10.22%**	4167	11988	0	255	1147	3596	311548
Services	**82.32%**	33559	124738	0	1847	5137	16930	3236282
Construction	**6.05%**	2464	8070	0	132	429	1452	153802
Natural Resources	**1.41%**	576	2543	0	76	181	413	67447

Notes: The table presents summary statistics that illustrate the sectoral distribution of county-level employment data for 2021. The second column indicates the percentage of national employment represented by each sector. The dataset includes a total of 3,079 counties.

The second column of the table indicates what is the percentage of the national employment that is represented by each of the sectors. It is noticeable that services composes the vast majority of the employment in the US, with a share exceeding 82%

in 2021. The remaining columns illustrate the distribution of the sectoral employment among the counties. Table 4.12 adds to the picture by showing how the sectoral shares are distributed among the counties.

Table 4.12: Summary statistics: distribution of county-level sectoral employment shares (2021)

Sector	Mean	Std. Dev.	Min	25%	Median	75%	Max
Manufacturing	0.151	0.124	0.000	0.056	0.124	0.217	0.904
Services	0.736	0.136	0.000	0.655	0.754	0.835	1.000
Construction	0.064	0.057	0.000	0.038	0.056	0.079	0.850
Natural Resources	0.050	0.077	0.000	0.006	0.021	0.058	0.807

Notes: The table provides summary statistics that illustrate the distribution of sectoral employment shares at the county level for the year 2021. The share for each county and sector is calculated by dividing the employment in that sector by the total employment in the county. The dataset includes a total of 3,079 counties.

Notice the mean shown in the table is not weighted by size of the employment, hence not matching with the share of the national employment displayed in the previous table. As the table suggests, the services sector is predominant in the majority of the counties. However, as can be seen by the last column, there are a few counties in which one of the other sectors might be the most important, but this phenomenon tend to occur in counties with a relatively low number of total workers. In a similar fashion as in Table 4.11, the table below illustrates the distribution of the sectoral expenditure.

Table 4.13: Summary statistics: distribution of sectoral expenditure

Sector	Share of national expenditure	Mean	Std. Dev.	25%	Median	75%	Max
Manufacturing	**76.08%**	146095965	1014526338	21661	585746	9376390	30542254977
Services	**22.00%**	42239148	317069279	12899	153796	1890427	11916529739
Construction	**1.89%**	3621675	130064220	0	34297	812381	829609240
Natural Resources	**0.03%**	69031	809779	0	0	2114	32971347

Notes: The table presents summary statistics that illustrate the sectoral distribution of county-level military procurement expenditure data. The data corresponds to the annual average expenditure over a range from 2006 to 2021. The second column shows the percentage of national expenditure attributed to each sector. The dataset includes a total of 3,079 counties.

Notice that the amounts shown in Table 4.13 refer to the average across all years in the dataset. The sectoral share of the expenditure differs substantially in comparison to the case of employment. Over 76% of the expenditure is allocated to the manufacturing sector, despite it representing only around 10% of the total employment. Meanwhile, services, the largest sector in terms of employment, receives 22% of the expenditure, while Construction receives next to 2% and Natural Resources receives a tiny share of 0.03%.

As also shown in the table, a considerable amount of counties receive no expenditure in some specific sectors, specially for Construction and Natural Resources. In fact, a lot of the expenditure in those two sectors are concentrated in very few counties, largely in the case of Natural Resources, as over half of the total counties has not received any expenditure in this sector for all the years of the data.

A set of equations is estimated in this section to capture different features of the sectoral effects of expenditure. As in the last section, dummy variables are used to separate counties given the level of expenditure. However, in contrast to the previous sections, the dummy variables used no longer indicate a high-expenditure percentile. Calculating such percentiles at the sectoral level is problematic due to the large number of zero or very low expenditure observations. For example, the share of county-sector-year observations with zero expenditure is 39% in manufacturing, 35% in services, 66% in construction, and 89% in natural resources. In the case of natural resources, the cutoff value for the top 33% of expenditure remains zero, preventing the meaningful division of observations into percentile-based groups. Addressing this issue would require either applying different percentile thresholds for each sector or defining a highly restrictive high-expenditure group to maintain consistency across all estimations.

To avoid these complications, I use a new dummy variable denoted $D^{pos}_{i,j,t}$, in which j indicates the sector. It is defined to take the value 1 when a county-sector pair receives any non-zero expenditure in a given year and 0 otherwise. This alternative specification ensures a consistent grouping criterion across sectors while accommodating the large proportion of zero-valued observations.

A dummy variable defined in such way leads to a key difference between the benchmark models (no dummy) and the ones that include the dummy variable. This difference lies in how areas with no variation in expenditure are treated. In the dummy-augmented models, the county-sector pairs with consistently zero expenditure contribute only to the estimation of the intercept but not to the slope. As a result, the slope coefficient is identified exclusively from observations where expenditure is positive. This design prevents the estimated slope from being disproportionately influenced by the substantial number of zero-expenditure observations.

Notice that all equations described below are estimated for all four sectors, which are indexed by j. The first two equations estimated are described below:

$$y_{i,j,t} = \theta_s + \gamma_t + \beta_0 exp_{i,t} + \mu_{i,j,t} \tag{4.7}$$

$$y_{i,j,t} = \theta_s + \gamma_t + \beta_0 D^{pos}_{i,t} + \beta_1 exp_{i,t} + \beta_2 D^{pos}_{i,t} exp_{i,t} + \mu_{i,j,t} \tag{4.8}$$

The goal of those equations is to examine how total (national) expenditure affects each sector individually. Notice the expenditure is not indexed to the sector level. The next two equations are analogous to (4.7) and (4.8), but instead they use the sectoral level of expenditure.

$$y_{i,j,t} = \theta_s + \gamma_t + \beta_0 exp_{i,j,t} + \mu_{i,j,t} \tag{4.9}$$

$$y_{i,j,t} = \theta_s + \gamma_t + \beta_0 D^{pos}_{i,j,t} + \beta_1 exp_{i,j,t} + \beta_2 D^{pos}_{i,j,t} exp_{i,j,t} + \mu_{i,j,t} \tag{4.10}$$

The following equations test for cross-sectoral effects of the expenditure. Equations (4.11) and (4.12) control for the effect of expenditure in the sector of interest plus the expenditure in one of the remaining sectors. For that reason, those equations are estimated repeated times switching the sector to which the cross effect is being estimated. Equations (4.13) and (4.14) are similar to the previous two equations, but they extend the model by controlling for the expenditure in all sectors at the same time.

$$y_{i,j,t} = \theta_s + \gamma_t + \beta_0\, exp_{i,j,t} + \beta_1\, exp_{i,k,t} + \mu_{i,j,t} \qquad k \neq j \tag{4.11}$$

$$\begin{aligned} y_{i,j,t} = \theta_s + \gamma_t + \beta_0 D^{pos}_{i,j,t} + \beta_1\, exp_{i,j,t} + \beta_2 D^{pos}_{i,j,t} exp_{i,j,t} + \\ \beta_3 D^{pos}_{i,k,t} + \beta_4\, exp_{i,k,t} + \beta_5 D^{pos}_{i,k,t} exp_{i,k,t} + \mu_{i,j,t} \qquad k \neq j \end{aligned} \tag{4.12}$$

$$y_{i,k,t} = \theta_s + \gamma_t + \sum_{j}^{J} \beta_j exp_{i,j,t} + \mu_{i,k,t} \qquad k \in J \tag{4.13}$$

$$y_{i,k,t} = \theta_s + \gamma_t + \sum_{j}^{J} \beta_0, jD_{i,j,t}^{pos} + \sum_{j}^{J} \beta_1, jexp_{i,j,t} + \sum_{j}^{J} \beta_2, jD_{i,j,t}^{pos} exp_{i,j,t} + \mu_{i,k,t} \qquad k \in J \tag{4.14}$$

Notice that J denotes the set containing all four sectors.

All the equations described below are estimated for each of the major sectors. Tables 4.14, 4.15, 4.16, and 4.17 presents the results for the employment in the manufacturing, services, construction, and natural resources as outcome variables, respectively.

Table 4.14: Estimation results for sectoral data Manufacturing

Specification	(4.7)	(4.8)	(4.9)	(4.10)	(4.11)			(4.12)			(4.13)	(4.14)
					Services	Construction	Natural Resources	Services	Construction	Natural Resources		
Expenditure (total)	0.157***	-0.129***										
	(0.005)	(0.014)										
D^{Pos}_{total} * Expenditure (total)		0.432***										
		(0.018)										
Expenditure (manufacturing)			0.153***	-0.132***	0.116***	0.139***	0.147***	-0.107***	-0.127***	-0.131***	0.112***	-0.107***
			(0.004)	(0.013)	(0.004)	(0.004)	(0.004)	(0.013)	(0.013)	(0.013)	(0.004)	(0.013)
D^{Pos}_{manuf} * Expenditure (manufacturing)				0.418***				0.313***	0.390***	0.404***		0.310***
				(0.017)				(0.017)	(0.017)	(0.017)		(0.017)
Expenditure (services)					0.074***			-0.058***			0.067***	-0.057***
					(0.004)			(0.015)			(0.004)	(0.015)
D^{Pos}_{serv} * Expenditure (services)								0.151***				0.147***
								(0.019)				(0.019)
Expenditure (construction)						0.036***			-0.035***		0.011***	-0.017*
						(0.004)			(0.010)		(0.004)	(0.009)
D^{Pos}_{cons} * Expenditure (construction)									0.058***			0.009
									(0.018)			(0.018)
Expenditure (natural resources)							0.047***			-0.072***	0.021***	-0.053***
							(0.006)			(0.016)	(0.006)	(0.016)
D^{Pos}_{NR} * Expenditure (natural resources)										0.072**		0.037
										(0.028)		(0.028)
D^{Pos}_{total}		-2.347***										
		(0.154)										
D^{Pos}_{manuf}			-2.140***					-1.411***	-1.908***	-1.996***		-1.398***
			(0.166)					(0.172)	(0.172)	(0.164)		(0.174)
D^{Pos}_{serv}								-0.396***				-0.375**
								(0.146)				(0.150)
D^{Pos}_{cons}									0.024			0.152
									(0.201)			(0.196)
D^{Pos}_{NR}										0.311		0.277
										(0.241)		(0.243)

Notes: The table displays the estimation results for equations (4.7) to (4.14), using employment in the manufacturing sector as the outcome variable. The first line of the table identifies the specific equation being estimated. Equations (4.11) and (4.12) were each estimated once for each of the remaining primary sectors, using their respective expenditures as independent variables. All variables in the regression have undergone an inverse hyperbolic sine transformation. The data covers the period from 2006 to 2021 and includes a total of 2,903 counties, resulting in 46,448 observations. All regressions account for both time and state fixed effects, and standard errors are clustered at the county level. The symbols "***", "**", and "*" indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 4.15: Estimation results for sectoral data
Services

Specification	(4.7)	(4.8)	(4.9)	(4.10)	(4.11)			(4.12)			(4.13)	(4.14)
					Manufacturing	Construction	Natural Resources	Manufacturing	Construction	Natural Resources		
Expenditure (total)	0.129***	-0.089***										
	(0.003)	(0.010)										
D^{Pos}_{total} * Expenditure (total)		0.386***										
		(0.013)										
Expenditure (manufacturing)					0.080***			-0.055***			0.073***	-0.054***
					(0.002)			(0.008)			(0.002)	(0.008)
D^{Pos}_{manuf} * Expenditure (manufacturing)								0.223***				0.216***
								(0.010)				(0.011)
Expenditure (services)			0.132***	-0.076***	0.083***	0.108***	0.122***	-0.050***	-0.072***	-0.075***	0.068***	-0.048***
			(0.003)	(0.010)	(0.003)	(0.003)	(0.003)	(0.009)	(0.009)	(0.010)	(0.002)	(0.009)
D^{Pos}_{serv} * Expenditure (services)				0.365***				0.213***	0.326***	0.347***		0.190***
				(0.012)				(0.012)	(0.013)	(0.012)		(0.012)
Expenditure (construction)						0.049***			-0.032***		0.029***	-0.020***
						(0.003)			(0.008)		(0.002)	(0.007)
D^{Pos}_{cons} * Expenditure (construction)									0.065***			0.028**
									(0.014)			(0.012)
Expenditure (natural resources)							0.065***			-0.074***	0.034***	-0.052***
							(0.005)			(0.013)	(0.005)	(0.011)
D^{Pos}_{NR} * Expenditure (natural resources)										0.079***		0.060***
										(0.024)		(0.021)
D^{Pos}_{total}		-2.770***										
		(0.098)										
D^{Pos}_{manuf}								-1.527***				-1.470***
								(0.111)				(0.112)
D^{Pos}_{serv}				-2.383***				-1.270***	-2.059***	-2.204***		-1.084***
				(0.097)				(0.089)	(0.098)	(0.093)		(0.087)
D^{Pos}_{cons}									-0.083			0.109
									(0.140)			(0.121)
D^{Pos}_{NR}										0.314		0.057
										(0.199)		(0.174)

Notes: The table displays the estimation results for equations (4.7) to (4.14), using employment in the services sector as the outcome variable. The first line of the table identifies the specific equation being estimated. Equations (4.11) and (4.12) were each estimated once for each of the remaining primary sectors, using their respective expenditures as independent variables. All variables in the regression have undergone an inverse hyperbolic sine transformation. The data covers the period from 2006 to 2021 and includes a total of 2,903 counties, resulting in 46,448 observations. All regressions account for both time and state fixed effects, and standard errors are clustered at the county level. The symbols "***", "**", and "*" indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 4.16: Estimation results for sectoral data
Construction

Specification	(4.7)	(4.8)	(4.9)	(4.10)	(4.11)			(4.12)			(4.13)	(4.14)
					Manufacturing	Services	Natural Resources	Manufacturing	Services	Natural Resources		
Expenditure (total)	0.151***	-0.100***										
	(0.005)	(0.016)										
D^{Pos}_{total} * Expenditure (total)		0.414***										
		(0.019)										
Expenditure (manufacturing)					0.122***			-0.084***			0.091***	-0.060***
					(0.004)			(0.019)			(0.004)	(0.018)
D^{Pos}_{manuf} * Expenditure (manufacturing)								0.331***				0.226***
								(0.022)				(0.021)
Expenditure (services)						0.123***			-0.094***		0.074***	-0.063***
						(0.004)			(0.016)		(0.004)	(0.015)
D^{Pos}_{serv} * Expenditure (services)									0.353***			0.197***
									(0.020)			(0.019)
Expenditure (construction)			0.111	-0.110***	0.056***	0.050***	0.097***	-0.054***	-0.046***	-0.100***	0.027***	-0.031***
			(0.004)	(0.013)	(0.004)	(0.004)	(0.004)	(0.010)	(0.010)	(0.013)	(0.003)	(0.009)
D^{Pos}_{cons} * Expenditure (construction)				0.296***				0.109***	0.075***	0.241***		0.037**
				(0.021)				(0.017)	(0.018)	(0.020)		(0.017)
Expenditure (natural resources)							0.077***			-0.127***	0.031***	-0.060***
							(0.006)			(0.017)	(0.006)	(0.013)
D^{Pos}_{NR} * Expenditure (natural resources)										0.134***		0.055**
										(0.033)		(0.026)
D^{Pos}_{total}		-2.662***										
		(0.137)										
D^{Pos}_{manuf}								-1.976***				-1.207***
								(0.160)				(0.156)
D^{Pos}_{serv}									-1.885***			-0.856***
									(0.143)			(0.133)
D^{Pos}_{cons}				-1.061***				-0.153	0.006	-0.617***		0.152
				(0.220)				(0.177)	(0.191)	(0.209)		(0.172)
D^{Pos}_{NR}										0.780***		0.229
										(0.277)		(0.226)

Notes: The table displays the estimation results for equations (4.7) to (4.14), using employment in the construction sector as the outcome variable. The first line of the table identifies the specific equation being estimated. Equations (4.11) and (4.12) were each estimated once for each of the remaining primary sectors, using their respective expenditures as independent variables. All variables in the regression have undergone an inverse hyperbolic sine transformation. The data covers the period from 2006 to 2021 and includes a total of 2,903 counties, resulting in 46,448 observations. All regressions account for both time and state fixed effects, and standard errors are clustered at the county level. The symbols "***", "**", and "*" indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 4.17: Estimation results for sectoral data
Natural resources

Specification	(4.7)	(4.8)	(4.9)	(4.10)	(4.11)			(4.12)			(4.13)	(4.14)
					Manufacturing	Services	Construction	Manufacturing	Services	Construction		
Expenditure (total)	0.058*** (0.004)	-0.044*** (0.014)										
D^{Pos}_{total} * Expenditure (total)		0.147*** (0.017)										
Expenditure (manufacturing)					0.051*** (0.004)			-0.032* (0.017)			0.037*** (0.003)	-0.022 (0.016)
D^{Pos}_{manuf} * Expenditure (manufacturing)								0.120*** (0.019)				0.084*** (0.020)
Expenditure (services)						0.049*** (0.004)			-0.054*** (0.014)		0.024*** (0.003)	-0.039*** (0.013)
D^{Pos}_{serv} * Expenditure (services)									0.134*** (0.017)			0.061*** (0.018)
Expenditure (construction)							0.034*** (0.003)			-0.049*** (0.010)	0.009*** (0.003)	-0.025*** (0.009)
D^{Pos}_{cons} * Expenditure (construction)										0.088*** (0.017)		0.025 (0.017)
Expenditure (natural resources)			0.054*** (0.006)	-0.079*** (0.019)	0.025*** (0.006)	0.026*** (0.006)	0.030*** (0.006)	-0.039** (0.019)	-0.040** (0.020)	-0.052*** (0.019)	0.013** (0.005)	-0.030 (0.019)
D^{Pos}_{NR} * Expenditure (natural resources)				0.119*** (0.033)				0.059* (0.031)	0.052 (0.032)	0.066** (0.032)		0.045 (0.031)
D^{Pos}_{total}		-0.715*** (0.134)										
D^{Pos}_{manuf}								-0.584*** (0.147)				-0.390** (0.154)
D^{Pos}_{serv}									-0.445*** (0.131)			0.019 (0.136)
D^{Pos}_{cons}										-0.051 (0.190)		0.109 (0.185)
D^{Pos}_{NR}				0.190 (0.243)				-0.016 (0.229)	0.091 (0.235)	0.177 (0.239)		-0.046 (0.230)

Notes: The table displays the estimation results for equations (4.7) to (4.14), using employment in the natural resources sector as the outcome variable. The first line of the table identifies the specific equation being estimated. Equations (4.11) and (4.12) were each estimated once for each of the remaining primary sectors, using their respective expenditures as independent variables. All variables in the regression have undergone an inverse hyperbolic sine transformation. The data covers the period from 2006 to 2021 and includes a total of 2,903 counties, resulting in 46,448 observations. All regressions account for both time and state fixed effects, and standard errors are clustered at the county level. The symbols "***", "**", and "*" indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

As shown in the estimation results of equations (4.7) and (4.8) in the tables above, the association between total procurement expenditure and sectoral employment differs substantially across sectors. These differences reflect both the characteristics of each sector and the differences in the base values of employment and expenditure reported in Tables 4.11, 4.12, and 4.13. The estimated elasticities should therefore not be directly compared across sectors, since a given elasticity may correspond to very different absolute changes in employment depending on the size of the sector. However, the results indicate that the employment response is substantially weaker in natural resources than in the other sectors, which is consistent with the very small share of total procurement expenditure allocated to this sector.

Equations (4.9) to (4.14) provide further insight into the relationship between sector-specific procurement expenditure and employment within the same sector. The estimated own-sector effects for manufacturing and services are substantially larger than those for construction and natural resources. In specifications that distinguish between high and low expenditure counties, the effect of manufacturing expenditure on manufacturing employment remains around 0.3 when services expenditure is not controlled for, and decreases to approximately 0.2 once services expenditure is included. A similar pattern is observed in the services sector, where the own-sector association ranges between approximately 0.25 and 0.3 when manufacturing expenditure is excluded, and falls to around 0.15 when it is included. By contrast, the estimated effects of construction and natural-resource expenditure on employment in their own sectors decline substantially once manufacturing and services expenditure are controlled for.

These patterns suggest that sectoral composition matters for the transmission of procurement expenditure. The stronger own-sector results for manufacturing and services are consistent with a direct production-channel mechanism: procurement raises demand for the goods and services produced by firms in those sectors, which in turn is associated with higher sectoral employment. On the other hand, the weaker own-sector effects in construction and natural resources are also intuitive. Construction procurement is often project-based and temporary, and may be fulfilled partly by mobile contractors or workers commuting from neighbouring areas. Natural-resource procurement represents a very small share of total defence expenditure and is likely to be more capital intensive or rely on some commodities already on inventory, so additional expenditure may translate less directly into local employment.

These findings also suggest that a substantial portion of the sectoral effects of procurement expenditure on employment arises from cross-sectoral spillovers. Procurement expenditure recorded in one sector does not necessarily impacts employment only in that same sector. A manufacturing contract, for example, may increase demand for logistics, maintenance, technical support, business services, and other supplier activities. Similarly, services procurement may support manufacturing employment when it is tied to the production of defence goods. For example, engineering, equipment testing, logistics, software integration, or technical support contracts may increase demand for manufacturing workers if they are part of a larger project involving aircraft, vehicles, electronics, weapons systems, or other defence equipment. In this sense, the sector in which procurement expenditure is recorded may not be the only sector in which employment responds.

This mechanism is supported by the estimates from equation (4.14). In high-expenditure counties, manufacturing expenditure is associated with spillover effects of 0.162, 0.166, and 0.062 on employment in services, construction, and natural resources, respectively. Services expenditure also generates positive spillovers, with estimated effects of 0.090, 0.134, and 0.022 on employment in manufacturing, construction, and natural resources, respectively. These estimates suggest that manufacturing and services play a central role in the local transmission of defence procurement, not only through their own employment responses but also through their links with other sectors.

The evidence for construction and natural resources is different. As already mentioned above, once manufacturing and services expenditure are included in the specification, the direct effects of construction and natural-resource procurement on employment in their own sectors become considerably weaker. This suggests that much of the employment association observed in these sectors may be driven by spillovers from manufacturing and services rather than by direct procurement effects within construction or natural resources themselves. This interpretation is consistent with the small share of total procurement expenditure allocated to construction and natural resources.

The stronger spillover effects from manufacturing and services can be explained by the role these sectors play in defence procurement supply chains. Manufacturing contracts often involve complex production processes that require inputs from several other activities, which includes elements such as logistics, technical services, construction-related support, and natural resource inputs such as fuel and minerals. As a result, an increase in manufacturing procurement can generate strong employment responses outside the manufacturing sector itself. A similar reasoning can be applied to the services sector, as the sector might require inputs from other sectors such as equipment, and vehicles.

In addition to supply-chain linkages, manufacturing and services may generate stronger spillovers through broader local demand effects. Because these sectors account for a large share of procurement expenditure, increases in activity in manufacturing and services can generate additional jobs and income within the county. This may raise demand for local goods and services, including food, retail, housing, transport, healthcare, and personal services. These local demand effects can support employment in sectors that are not directly involved in defence procurement. In this sense, the spillovers from manufacturing and services may reflect not only production linkages, but also broader general-equilibrium effects operating through household income, local consumption, housing demand, and firm entry or expansion in the local economy.

The positive cross-sectoral spillovers are also consistent with the broader labour-market results presented earlier in the chapter. A possible concern is that procurement expenditure in one sector could draw workers away from other sectors within the same county, generating negative employment effects elsewhere. The results provide limited evidence of this type of substitution. Instead, the positive spillover coefficients suggest that procurement-intensive sectors may expand local labour demand in a way that supports employment in other parts of the county economy. This is consistent with the interpretation that local adjustment occurs through a combination of commuting, migration, supplier linkages, and local demand effects, rather than solely through the reallocation of existing workers across sectors.

It is important to note a limitation of the approach used in this section. Given the specifications employed, the estimated sectoral spillover effects are confined to within-county dynamics and do not capture any cross-geographical spillovers. For instance, manufacturing procurement in one county may increase demand for inputs, services, or logistics provided by firms in other counties. In that case, some employment effects would occur outside the county where the procurement expenditure is recorded. As a result, the elasticities reported here may underestimate the full extent of sectoral spillover effects, in particular those that operate through interregional production networks and supply chains.

Finally, the interpretation of the sectoral results should be read alongside the fixed-effect discussion in the previous section. The results are informative about how procurement exposure is associated with the sectoral structure of local employment, but they should not be interpreted as evidence of large short-run within-county employment responses to annual procurement changes. Rather, the patterns are consistent with the idea that counties with stronger defence-related manufacturing and specialised service bases are also the places where procurement expenditure is most strongly associated with employment. This reinforces the broader conclusion that sectoral composition is central to understanding the local labour-market impacts of defence procurement.

4.5 Concluding remarks

This chapter has examined the regional and sectoral effects of military procurement expenditure on local labour markets in the US counties, with particular attention to labour mobility, spatial spillovers and cross-sectoral dynamics. The analysis revealed that procurement expenditure positively affects local employment, labour force participation, and population levels, with particularly pronounced effects in counties with high expenditure levels. The findings indicate a notably strong relationship between spending and population growth, suggesting that relocation of workers plays a key role in filling new job positions through the migration of individuals.

The estimated elasticities reveal that for an increase of 1% in expenditure employment increases by 0.42%, the labour force by 0.38% and population by 0.36%. These results should, however, be interpreted alongside the additional specifications estimated with county fixed effects. When county fixed effects are included, the estimated coefficients for employment, labour force, and population become substantially smaller and are generally close to zero. This indicates that the main results are not primarily driven by short-run changes in procurement expenditure within the same county over time. Rather, they appear to reflect persistent differences in procurement exposure across counties.

The results also suggest the presence of spatial spillover effects, with increases in procurement expenditure within a county positively impacting labour market indicators in neighbouring counties that are part of the same local labour market. This effect is likely driven by indirect channels, such as income effects, where higher local incomes stimulate demand for goods and services in adjacent areas, and demand for inputs, as firms in neighbouring counties may supply goods or services to the county with heightened expenditure. Notably, these positive spillover channels appear strong enough to outweigh potential job-leeching effects, where workers might otherwise leave neighbouring counties to take advantage of new job opportunities in the area receiving the expenditure shock.

Comparing spillover effects across CMZs and LMAs further highlights the importance of economic linkages. The magnitudes of spillovers are consistently larger in CMZs, suggesting that those labour markets exhibit tighter integration than those delineated by broader administrative or population-based criteria, such as LMAs. The stronger interdependencies within CMZs likely amplify positive spillovers, as workers, firms, and supply chains are more closely connected within these regions. For instance, an increase of 1% in the CMZ a county belongs to leads to a 0.08% increase on the counties employment and 0.09% in population, while for LMA, the increases are 0.06% and 0.06%, respectively.

When examining the effects of procurement expenditure on wages, the estimated impacts appear relatively modest. In counties with high expenditure, a 1% increase in local spending corresponds to only a 0.05% increase in local wages. However, large swings in expenditure can accumulate to produce significantly greater effects on wage levels. Such episodes of substantial expenditure increases occurred with some regularity over the study period.

Evidence of spillover effects is also observed: a 1% increase in expenditure in neighboring counties within the same local labor market is associated with an approximate 0.02% increase in local wages. These findings align with earlier observations on migration and labor mobility. When rising labor demand from increased expenditure is met by an influx of workers, the resulting upward pressure on local wages may be mitigated. This chapter also investigates potential asymmetries in wage responses to expenditure shocks, motivated by the hypothesis of downward wage rigidity. However, the analysis finds no strong evidence supporting the existence of such asymmetries. Finally, a limitation of the wage analysis is its reliance on nominal wages, which may overstate real wage effects due to the lack of county-level price data.

A natural concern with the results in this chapter is the exogeneity of military procurement expenditure. Although defence procurement is plausibly more exogenous than many other forms of public spending, it can still be endogenous to local conditions: contracts are more likely to be placed where contractor capacity, skilled labour, and political influence are already present. This can make expenditure positively correlated with unobserved local demand, biasing OLS estimates upward. To strengthen identification, a straightforward next step is to construct a shift–share instrument using pre-period sectoral shares interacted with national growth in DoD procurement by sector.

This chapter also provides evidence of heterogeneity in the sectoral effects of procurement expenditure on labour markets, as well as strong cross-sectoral spillover dynamics. The manufacturing sector accounts for the vast majority of total DoD procurement expenditure, exceeding 76%. It is also responsible for substantial spillover effects across all other sectors. The services sector likewise produces notable spillovers, with coefficients not far below those for manufacturing, an especially striking result given its much smaller share of total expenditure.

By contrast, procurement directed at the construction and natural resources sectors represents only a small share of total expenditure. The most striking finding is that spending in these sectors produces little to no observable spillover effects on labour markets. Indeed, when controlling for manufacturing and services expenditure, even the direct impacts of construction and natural resources spending on their own sectoral employment virtually disappear. This suggests that most employment effects in these sectors are driven by indirect demand spillovers originating in manufacturing and services, rather than by direct procurement targeted to them.

Chapter 5

Conclusion

This thesis presented three essays that turn to regional data to explore various aspects of local and national labour market dynamics and the effects of fiscal policy. Chapter 2 estimates a county-level wage Phillips curve with explicit spatial spillovers arising from labour mobility, clarifying how labour market tightness translates into wage growth once commuting patterns are accounted for. Chapter 3 shifts to data infrastructure, documenting a highly granular dataset of military procurement and providing tools for its systematic use; this dataset also bridges to Chapter 2, since procurement shocks plausibly raise local labour demand, tighten markets, and thereby affect wages. Chapter 4 makes this link explicit by using the data from Chapter 3 to quantify the effects of military procurement on employment, wages, and population, and to trace both spatial and cross-sectoral spillovers, reconnecting with the mobility mechanisms highlighted in Chapter 2. Together, the essays establish a coherent regional-macro perspective and open avenues for future research.

The results from Chapter 2 indicate a steep national wage Phillips curve, with coefficients (in absolute value) exceeding unity and ranging from -1.063 to -1.657 depending on the choice of neighbour matrix. Interpreted under a slack measure (e.g., unemployment), these magnitudes imply that tighter labour markets translate strongly into faster wage growth, an important consideration for wage–price spiral risks and the inflation–unemployment trade-off. The results also point to a flattening curve after the 2008 crisis. Beyond the aggregate slope, the chapter highlights mechanisms of direct policy relevance: a sizeable share of the aggregate effect of labour market shocks on wage inflation reflects spillovers that extend beyond the initially affected county. To put the coefficients mentioned above into perspective, spillover effects explain an average share of 36% of the impacts of unemployment on national wage inflation. Moreover, idiosyncratic regional shocks have different national implications depending on each county's size-adjusted centrality: populous, highly connected counties, such as metropolitan cores, propagate shocks more widely and therefore warrant particular attention in policy design.

The chapter also shows that Labour Market Areas (LMAs) provide a valuable basis for constructing neighbour matrices in spatial wage models; Commuting Zones (CMZs) are likewise viable, but LMAs perform better for the dynamics studied here, partly because of their larger functional regions. A practical limitation is that LMAs were discontinued after 1990, so the underlying commuting inputs are dated; a natural path for future work is to recreate LMA data by applying Tolbert and Sizer's (1996) methodology with more recent Census flows. A deeper limitation of both LMAs and CMZs is their hard partitioning: by construction, each county belongs to a single zone, which can obscure cross-market ties (including between adjacent counties assigned to different clusters). Methodologically, an improvement would be to replace hard clusters with flow-based or overlapping neighbour definitions, for example, continuous weights based on inter-county commuting shares, or multi-membership spatial matrices, so that each county has its own empirically grounded set of neighbours.

Chapter 3 documents in detail a military expenditure dataset compiled from U.S. Department of Defence (DoD) procurement records, designed to be used in a multitude of studies regarding fiscal policy. The chapter highlights the dataset's rich granularity, with expenditure mapped at the county level (geography) and by sector, and, to a lesser extent, by firm. A key feature of the DoD procurement expenditure discussed in the chapter is the distribution of the values across the procurement records. A vast majority of the total expenditure is represented by a small number of contracts, showing that most of the volume of expenditure is allocated to a small number of companies in the form of high-value and long-length contracts. Also, the expenditure is highly concentrated in a limited number of sectors, with predominance of the ones related to combat vehicles and weapons. The chapter also addresses the treatment of contract length in aggregation. When contract values are distributed over their active months prior to aggregation, the resulting series becomes more persistent and less noisy, especially at higher frequencies. This differentiation is also more impactful the higher the level of granularity, as fewer procurements are observed.

Chapter 4 applies the procurement dataset from the preceding chapter to estimate elasticities of DoD spending with respect to local labour market outcomes. A central finding is that higher procurement expenditure induces worker reallocation across counties, with migration helping to fill newly created vacancies. Modest geographical spillovers are also detected. Exploiting the sectoral dimension, the chapter documents heterogeneous sector responses and identifies cross-sectoral spillovers, with manufacturing and services acting as key conduits that transmit shocks to the broader local economy.

The chapter also links back to Chapter 2 on wage determination. Whereas Chapter 2 analyses the relationship between unemployment and wages in a spatially augmented wage Phillips curve, Chapter 4 examines how local demand shocks, in the form of procurement expenditure, affect employment and wages, thereby complementing the earlier evidence on wage–slack dynamics with a spending-based perspective. One limitation of the chapter and a scope for future work is addressing the potential endogeneity in DoD procurement allocations through the use of shift-share instruments.

Bibliography

Acemoglu, D., U. Akcigit, and W. Kerr (2016). Networks and the macroeconomy: An empirical exploration. *Nber macroeconomics annual 30*(1), 273–335.

Adelman, M. A. (1969). Comment on the" h" concentration measure as a numbers-equivalent. *The Review of economics and statistics*, 99–101.

Auerbach, A., Y. Gorodnichenko, P. B. McCrory, and D. Murphy (2022). Fiscal multipliers in the covid19 recession. *Journal of International Money and Finance 126*, 102669.

Auerbach, A., Y. Gorodnichenko, and D. Murphy (2020). Local fiscal multipliers and fiscal spillovers in the usa. *IMF Economic Review 68*(1), 195–229.

Babb, N. and A. K. Detmeister (2017). Nonlinearities in the phillips curve for the united states: Evidence using metropolitan data.

Barattieri, A., M. Cacciatore, and N. Traum (2023). Estimating the effects of government spending through the production network.

Bellemare, M. F. and C. J. Wichman (2020). Elasticities and the inverse hyperbolic sine transformation. *Oxford Bulletin of Economics and Statistics 82*(1), 50–61.

Beraja, M., E. Hurst, and J. Ospina (2019). The aggregate implications of regional business cycles. *Econometrica 87*(6), 1789–1833.

Bishop, J., E. Greenland, et al. (2021). *Is the Phillips Curve Still a Curve?: Evidence from the Regions.* Reserve Bank of Australia.

Blanchard, O. and R. Perotti (2002). An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *the Quarterly Journal of economics 117*(4), 1329–1368.

Borusyak, K., P. Hull, and X. Jaravel (2025). A practical guide to shift-share instruments. *Journal of Economic Perspectives 39*(1), 181–204.

Bouakez, H., O. Rachedi, and E. Santoro (2023). The government spending multiplier in a multisector economy. *American Economic Journal: Macroeconomics 15*(1), 209–239.

Burya, A., R. Mano, Y. Timmer, and A. Weber (2023, May). The wage phillips curve under labor market power. *AEA Papers and Proceedings 113*, 110–13.

Chodorow-Reich, G. (2019). Geographic cross-sectional fiscal spending multipliers: What have we learned? *American Economic Journal: Economic Policy 11*(2), 1–34.

Clemens, J. and S. Miran (2012). Fiscal policy multipliers on subnational government spending. *American Economic Journal: Economic Policy 4*(2), 46–68.

Cox, L., G. J. Müller, E. Pasten, R. Schoenle, and M. Weber (2024). Big g. *Journal of Political Economy 132*(10), 3260–3297.

Demyanyk, Y., E. Loutskina, and D. Murphy (2019). Fiscal stimulus and consumer debt. *Review of Economics and Statistics 101*(4), 728–741.

Dupor, B. and R. Guerrero (2017). Local and aggregate fiscal policy multipliers. *Journal of Monetary Economics 92*, 16–30.

Dupor, B. and P. B. McCrory (2018). A cup runneth over: Fiscal policy spillovers from the 2009 recovery act. *The Economic Journal 128*(611), 1476–1508.

Fisher, J. D. and R. Peters (2010). Using stock returns to identify government spending shocks. *The Economic Journal 120*(544), 414–436.

Fitzgerald, T., C. Jones, M. Kulish, and J. P. Nicolini (2024). Is there a stable relationship between unemployment and future inflation? *American Economic Journal: Macroeconomics 16*(4), 114–142.

Foote, A., M. J. Kutzbach, and L. Vilhuber (2021). Recalculating...: How uncertainty in local labour market definitions affects empirical findings. *Applied Economics 53*(14), 1598–1612.

Galí, J. and L. Gambetti (2019). Has the us wage phillips curve flattened? a semi-structural exploration. Technical report, National Bureau of Economic Research.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020). Bartik instruments: What, when, why, and how. *American Economic Review 110*(8), 2586–2624.

Hall, R. E. (2009). By how much does gdp rise if the government buys more output? Technical report, National Bureau of Economic Research.

Hazell, J., J. Herreno, E. Nakamura, and J. Steinsson (2022). The slope of the phillips curve: evidence from us states. *The Quarterly Journal of Economics 137*(3), 1299–1344.

Hooker, M. and M. M. Knetter (1994). Unemployment effects of military spending: evidence from a panel of states.

Hooker, M. A. and M. M. Knetter (1997). The effects of military spending on economic activity: Evidence from state procurement spending. *Journal of Money, Credit, and Banking*, 400–421.

Hooper, P., F. S. Mishkin, and A. Sufi (2020). Prospects for inflation in a high pressure economy: Is the phillips curve dead or is it just hibernating? *Research in Economics 74*(1), 26–62.

Kumar, A. and P. M. Orrenius (2016). A closer look at the phillips curve using state-level data. *Journal of Macroeconomics 47*, 84–102.

Leduc, S., D. J. Wilson, et al. (2017). Has the wage phillips curve gone dormant? *FRBSF Economic Letter 30*, 16.

McLeay, M. and S. Tenreyro (2020). Optimal inflation and the identification of the phillips curve. *NBER Macroeconomics Annual 34*(1), 199–255.

Mintz, A. (2002). *The political economy of military spending in the United States*. Routledge.

Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika 37*(1/2), 17–23.

Nakamura, E. and J. Steinsson (2014). Fiscal stimulus in a monetary union: Evidence from us regions. *American Economic Review 104*(3), 753–792.

Oatley, T. (2015). *A Political Economy of American Hegemony*. Cambridge University Press.

Ramey, V. A. (2011a). Can government purchases stimulate the economy? *Journal of Economic Literature 49*(3), 673–685.

Ramey, V. A. (2011b). Identifying government spending shocks: It's all in the timing. *The Quarterly Journal of Economics 126*(1), 1–50.

Ramey, V. A. and M. D. Shapiro (1998). Costly capital reallocation and the effects of government spending. In *Carnegie-Rochester conference series on public policy*, Volume 48, pp. 145–194. Elsevier.

Smith, C. (2014). The effect of labor slack on wages: Evidence from state-level relationships. Technical report, Board of Governors of the Federal Reserve System (US).

Tiebout, C. M. (1956). A pure theory of local expenditures. *Journal of political economy 64*(5), 416–424.

Tolbert, C. M. and M. Sizer (1996). Us commuting zones and labor market areas: A 1990 update.

Appendix A

Alternative neighbouring matrices

In this appendix, I re-estimate specifications (II) and (III) using neighbouring matrices with a higher level of complexity than a simple contiguity matrix as a robustness check. I test two different classes of matrices: a fixed radius matrix and an inverse distance matrix. Both rely on the concept of a centroid. A county's centroid is the point representing its geometric centre. It is computed by taking the county's average coordinates.

Fixed radius matrix. This matrix considers as neighbours the counties whose centroids are within a predetermined distance. I test using four different radius sizes[1]: 150km, 300km, 500km, and 1000km.

Inverse distance matrix. This matrix uses the inverse distance between the centroids of each pair of regions as weights. Similar to the fixed radius matrix approach, the possible pairings are restricted to a specified radius, with all regions outside this radius assigned a value of 0. The weight of each neighbour is determined as $1/distance$. Additionally, all rows are normalized to ensure that they sum to 1[2].

Tables A1 and A3 present the results for the fixed radius matrix and the inverse distance matrix. Tables A2 and A4 display the same results but for the IV estimation.

[1] The predetermined distance cannot be too small, as some counties have no neighbours if a small radius is used.

[2] This is in essence the same as the fixed radius matrix, but the matrix attributes weights accordingly to the distance.

Table A1: Fixed radius W matrix OLS estimation results

Radius	150km		300km		500km		1000km	
Specifiction	**(II)**	**(III)**	**(II)**	**(III)**	**(II)**	**(III)**	**(II)**	**(III)**
Δw_{t-1}	-0.139***	-0.139***	-0.138***	-0.138***	-0.137***	-0.137***	-0.133***	-0.133***
	(0.014)	(0.003)	(0.014)	(0.003)	(0.014)	(0.003)	(0.014)	(0.003)
$W * \Delta w_{t-1}$	0.255***	0.190***	0.391***	0.286***	0.482***	0.386***	0.613***	0.391***
	(0.018)	(0.015)	(0.026)	(0.029)	(0.033)	(0.046)	(0.052)	(0.071)
U_t	-0.179***	-0.177***	-0.216***	-0.212***	-0.235***	-0.236***	-0.240***	-0.234***
	(0.021)	(0.012)	(0.020)	(0.011)	(0.019)	(0.011)	(0.018)	(0.010)
$W * U_t$	-0.120***	-0.152***	-0.017	-0.078**	0.055*	0.044	0.073*	-0.114
	(0.025)	(0.024)	(0.026)	(0.038)	(0.029)	(0.054)	(0.043)	(0.070)
$W * \mu_t$		0.346***		0.555***		0.693***		0.797***
		(0.009)		(0.013)		(0.015)		(0.016)

Notes: Results for the estimation of equations II, and III. Each pair of columns represent the estimation for a given radius used to compute the neighbouring matrix W. The data range is 1992-2021. The standard errors were clustered at the county level.
"***", "**" and "*" denote statistical significance at the 1, 5 and 10 per cent levels.

Table A2: Fixed radius W matrix: Bartik-style instruments

Radius	150km		300km		500km		1000km	
Specifiction	**(II)**	**(III)**	**(II)**	**(III)**	**(II)**	**(III)**	**(II)**	**(III)**
				Second Stage				
Δw_{t-1}	0.155***	-0.162***	-0.149***	-0.157***	-0.147***	-0.154***	-0.145***	-0.150***
	(0.014)	(0.004)	(0.015)	(0.004)	(0.015)	(0.004)	(0.016)	(0.004)
$W * \Delta w_{t-1}$	0.019	-0.097***	0.109*	-0.323***	0.331***	-0.402***	0.310**	-0.626***
	(0.034)	(0.027)	(0.062)	(0.059)	(0.085)	(0.090)	(0.123)	(0.129)
U_t	-0.484***	-0.673***	-0.726***	-0.921***	-0.875***	-1.067***	-0.889***	-1.021***
	(0.158)	(0.089)	(0.162)	(0.092)	(0.161)	(0.091)	(0.159)	(0.090)
$W * U_t$	-1.249***	-1.595***	-0.397**	-1.631***	0.418*	-1.386***	0.407**	-1.407***
	(0.162)	(0.122)	(0.185)	(0.198)	(0.223)	(0.259)	(0.184)	(0.278)
$W * \mu_t$		0.364***		0.580***		0.718***		0.822***
		(0.009)		(0.013)		(0.015)		(0.015)
				First Stage				
Δw_{t-1}	-0.016***		-0.015***		-0.016***		-0.018***	
	(0.001)		(0.001)		(0.001)		(0.001)	
$W * \Delta w_{t-1}$	-0.194***		-0.322***		-0.416***		-0.675***	
	(0.004)		(0.006)		(0.007)		(0.012)	
Bartik	-0.240***		-0.233***		-0.234***		-0.231***	
	(0.007)		(0.007)		(0.007)		(0.007)	

Notes: Results for the IV estimation of equations II, and III. The W matrix was computed use an inverse distance matrix. Each pair of columns represent the estimation for a radius used to limit the neighbourhood. The top part of the table reports the results of the second stage of the IV procedure, while the bottom part reports the first stage. The data range is 1992-2021. The standard errors were clustered at the county level.
"***", "**" and "*" denote statistical significance at the 1, 5 and 10 per cent levels.

Table A3: Inverse distance W matrix: estimation results

Radius	**500km**		**1000km**	
Specifiction	**(II)**	**(III)**	**(II)**	**(III)**
Δw_{t-1}	-0.141***	-0.141***	-0.139***	-0.140***
	(0.014)	(0.003)	(0.014)	(0.003)
$W * \Delta w_{t-1}$	0.500***	0.541***	0.672***	0.869***
	(0.032)	(0.040)	(0.046)	(0.061)
U_t	-0.215***	-0.199***	-0.224***	-0.201***
	(0.021)	(0.012)	(0.020)	(0.012)
$W * U_t$	0.012	-0.103**	0.061	-0.124*
	(0.030)	(0.052)	(0.039)	(0.074)
$W * \mu_t$		0.693***		0.858***
		(0.013)		(0.011)

Notes: Results for the two-stage maximum likelihood estimation of equations II, and III. Each pair of columns represent the estimation for a given radius used to compute the neighbouring matrix W. The data range is 1992-2021. The standard errors were clustered at the county level.
"***", "**" and "*" denote statistical significance at the 1, 5 and 10 per cent levels.

Table A4: Inverse distance W matrix: Bartik-style instruments

Radius	**500km**		**1000km**	
Specifiction	**(II)**	**(III)**	**(II)**	**(III)**
		Second Stage		
Δw_{t-1}	-0.149***	-0.152***	-0.150***	-0.151***
	(0.015)	(0.004)	(0.015)	(0.004)
$W * \Delta w_{t-1}$	0.578***	0.324***	1.004***	0.812***
	(0.086)	(0.081)	(0.131)	(0.114)
U_t	-0.962***	-1.011***	-1.008***	-1.038***
	(0.166)	(0.095)	(0.166)	(0.096)
$W * U_t$	1.146***	0.404*	1.911***	1.444***
	(0.212)	(0.242)	(0.218)	(0.277)
$W * \mu_t$		-0.702***		0.856***
		(0.013)		(0.011)
		First Stage		
Δw_{t-1}	-0.013***		-0.014***	
	(0.001)		(0.001)	
$W * \Delta w_{t-1}$	-0.403***		-0.603***	
	(0.006)		(0.009)	
Bartik	-0.226***		-0.222***	
	(0.007)		(0.007)	

Notes: Results for the two-stage maximum likelihood estimation of equations II, and III. Each pair of columns represent the estimation for a given radius used to compute the neighbouring matrix W. The top part of the table reports the results of the second stage of the IV procedure, while the bottom part reports the first stage. The data range is 1992-2021. The standard errors were clustered at the county level.
"***", "**" and "*" denote statistical significance at the 1, 5 and 10 per cent levels.

The results presented above indicate considerable volatility as the predetermined distances increase. Notably, the magnitudes of the coefficients on $W * \Delta w_{t-1}$ and $W * \mu_t$ show significant growth with the radius size. Additionally, the coefficient on $W * U_t$ changes sign in the model (II) and loses significance in the model (III) as the distance increases in some specifications.

Also, the time fixed effects change substantially between each estimation. Tables A5 and A6 display the time fixed effects for the fixed radius estimations from Tables A1 and A2.

Table A5: Time fixed effects: OLS/ML estimations

Year/Radius	150km		300km		500km		1000km	
Specification	**(2)**	**(3)**	**(2)**	**(3)**	**(2)**	**(3)**	**(2)**	**(3)**
1992	0.059	0.017	0.047	0.016	0.038	0.014	0.032	0.016
1993	0.034	-0.008	0.021	-0.009	0.012	-0.011	0.006	-0.010
1994	0.042	-0.001	0.033	0.000	0.027	0.000	0.023	0.001
1995	0.041	-0.002	0.031	-0.002	0.024	-0.002	0.019	-0.002
1996	0.045	0.003	0.035	0.003	0.029	0.003	0.024	0.003
1997	0.051	0.009	0.041	0.009	0.034	0.009	0.028	0.008
1998	0.046	0.004	0.035	0.003	0.027	0.003	0.021	0.002
1999	0.039	-0.004	0.028	-0.004	0.021	-0.004	0.015	-0.006
2000	0.044	0.001	0.035	0.002	0.029	0.003	0.024	0.001
2001	0.038	-0.004	0.028	-0.004	0.021	-0.004	0.016	-0.005
2002	0.038	-0.005	0.028	-0.005	0.021	-0.004	0.016	-0.004
2003	0.040	-0.003	0.030	-0.003	0.023	-0.002	0.019	-0.002
2004	0.048	0.005	0.038	0.006	0.032	0.006	0.028	0.006
2005	0.040	-0.002	0.030	-0.002	0.022	-0.002	0.017	-0.003
2006	0.050	0.007	0.040	0.007	0.034	0.008	0.029	0.007
2007	0.048	0.006	0.037	0.005	0.030	0.005	0.024	0.004
2008	0.046	0.004	0.034	0.003	0.026	0.002	0.020	0.002
2009	0.028	-0.013	0.015	-0.015	0.006	-0.018	0.000	-0.013
2010	0.048	0.005	0.038	0.006	0.032	0.005	0.029	0.010
2011	0.047	0.005	0.036	0.005	0.028	0.003	0.024	0.007
2012	0.042	-0.001	0.031	-0.001	0.023	-0.002	0.019	0.001
2013	0.032	-0.011	0.021	-0.011	0.014	-0.011	0.010	-0.009
2014	0.043	0.000	0.035	0.001	0.029	0.003	0.027	0.003
2015	0.034	-0.009	0.024	-0.009	0.017	-0.008	0.013	-0.009
2016	0.020	-0.023	0.011	-0.022	0.005	-0.021	0.002	-0.022
2017	0.039	-0.006	0.032	-0.003	0.028	0.000	0.026	-0.002
2018	0.037	-0.006	0.028	-0.005	0.022	-0.003	0.018	-0.006
2019	0.034	-0.009	0.025	-0.009	0.019	-0.007	0.014	-0.010
2020	0.076	0.033	0.065	0.033	0.058	0.033	0.054	0.034
2021	0.047	0.006	0.034	0.004	0.025	0.002	0.016	0.000

Notes: Fixed effects from the estimation of equations (II) and (III) when fixed radius matrices are used as neighbouring matrices (W).
"***", "**" and "*" denote statistical significance at the 1, 5 and 10 per cent levels.

Table A6: Time fixed effects: two-stage IV estimation

Year/Radius	**150km**		**300km**		**500km**		**1000km**	
Specification	**(2)**	**(3)**	**(2)**	**(3)**	**(2)**	**(3)**	**(2)**	**(3)**
1992	0.193	0.049	0.123	0.055	0.055	0.054	0.059	0.055
1993	0.161	0.015	0.094	0.020	0.029	0.020	0.033	0.023
1994	0.153	0.000	0.093	-0.001	0.037	-0.002	0.041	-0.004
1995	0.150	-0.004	0.090	-0.004	0.035	-0.005	0.039	-0.005
1996	0.155	0.002	0.095	0.001	0.040	0.001	0.043	0.001
1997	0.156	0.001	0.098	0.000	0.045	0.001	0.048	0.002
1998	0.148	-0.008	0.092	-0.008	0.039	-0.006	0.042	-0.004
1999	0.135	-0.022	0.081	-0.023	0.031	-0.022	0.034	-0.020
2000	0.131	-0.030	0.081	-0.035	0.036	-0.033	0.039	-0.033
2001	0.137	-0.020	0.082	-0.021	0.031	-0.020	0.035	-0.018
2002	0.145	-0.009	0.086	-0.009	0.032	-0.009	0.036	-0.009
2003	0.150	-0.003	0.090	-0.004	0.035	-0.004	0.038	-0.005
2004	0.154	-0.001	0.096	-0.002	0.042	-0.002	0.046	-0.003
2005	0.145	-0.010	0.088	-0.010	0.034	-0.009	0.037	-0.007
2006	0.146	-0.013	0.092	-0.015	0.043	-0.014	0.046	-0.014
2007	0.145	-0.012	0.091	-0.012	0.040	-0.011	0.044	-0.008
2008	0.157	0.004	0.097	0.006	0.039	0.007	0.043	0.009
2009	0.184	0.049	0.105	0.059	0.027	0.056	0.032	0.057
2010	0.200	0.063	0.122	0.067	0.049	0.062	0.053	0.057
2011	0.194	0.055	0.119	0.061	0.046	0.058	0.051	0.055
2012	0.177	0.033	0.106	0.038	0.039	0.036	0.043	0.034
2013	0.160	0.013	0.092	0.016	0.029	0.014	0.033	0.012
2014	0.153	0.000	0.093	-0.003	0.039	-0.005	0.043	-0.008
2015	0.137	-0.019	0.080	-0.020	0.027	-0.020	0.031	-0.020
2016	0.117	-0.040	0.063	-0.044	0.014	-0.044	0.017	-0.046
2017	0.123	-0.039	0.074	-0.048	0.032	-0.048	0.035	-0.052
2018	0.120	-0.043	0.071	-0.049	0.028	-0.048	0.032	-0.047
2019	0.115	-0.049	0.067	-0.055	0.025	-0.053	0.028	-0.052
2020	0.196	0.047	0.132	0.049	0.072	0.048	0.075	0.047
2021	0.146	-0.010	0.091	-0.007	0.038	-0.004	0.042	0.003

Notes: Fixed effects from the two stage IV estimation of equations (II) and (III) using Bartik-style instruments when fixed radius matrices are used as neighbouring matrices (W).
"***", "**" and "*" denote statistical significance at the 1, 5 and 10 per cent levels.

The tables above present intriguing findings. First, a comparison of the results indicates that the time-fixed effects are, on average, greater for the two-stage estimation procedure. Second, the inclusion of spatial dependence within the error term reduces the magnitude of the time-fixed effects. This problem is in line with an issue from the spatial econometrics literature known as *spatial confounding*, which arises due to potential collinearity between the fixed effects and the spatial component in the error term. Lastly, expanding the radius size leads to a reduction in the time-fixed effects. This occurs because a larger distance encompasses more regions, transitioning the spatial lag of the variables from a localized focus to a national perspective. This shift is concerning, as the time-fixed effects are already accounting for variations at the national level, thereby amplifying the collinearity issue.

Given the results discussed above, I reach two conclusions. First, specification (II) is preferred over (III), as including spatially correlated error and fixed effects together might be problematic. Second, the matrix W should not include too many counties as neighbours; otherwise, it loses its property of measuring the local neighbourhood shocks and starts capturing aggregate-level features, which are already controlled by the time-fixed effects.

Appendix B

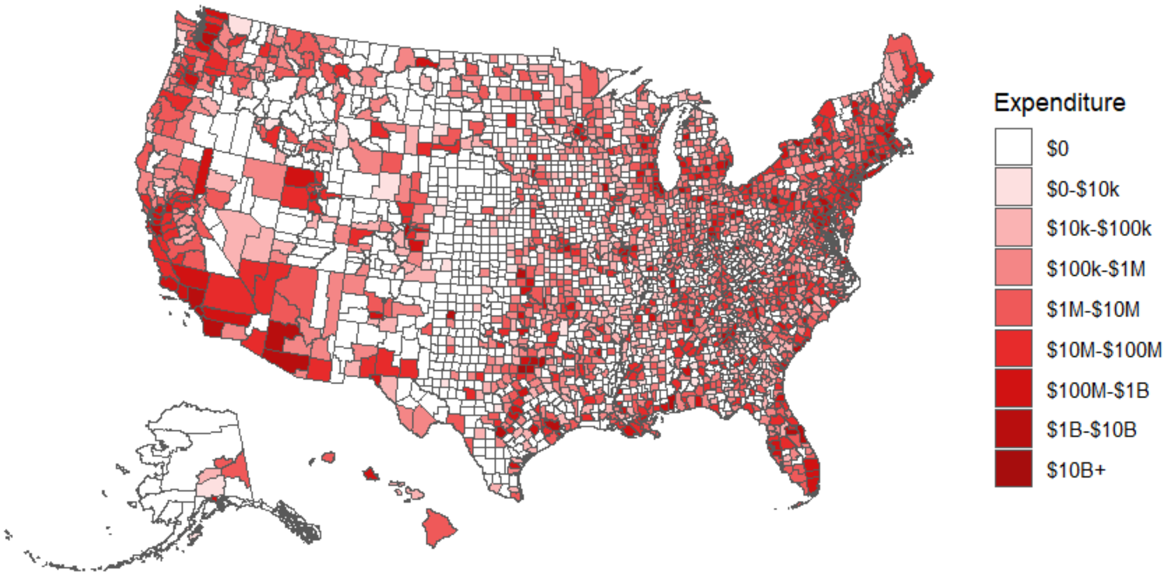
County-level distribution of manufacturing and PSTS DoD expenditure

Table B1: Top 20 counties by manufacturing DoD procurement expenditure for 2010, 2015, and 2020

2010		2015		2020	
County	**Total expenditure (millions of dollars) and share of total**	**County**	**Total expenditure (millions of dollars) and share of total**	**County**	**Total expenditure (millions of dollars) and share of total**
Tarrant, TX	11,022.54 (5.77%)	*Tarrant, TX*	9,783.91 (6.97%)	*Tarrant, TX*	24,523.19 (11.66%)
Los Angeles, CA	8,819.48 (4.62%)	*San Diego, CA*	5,528.02 (3.94%)	*Dallas, TX*	13,451.62 (6.40%)
Winnebago, WI	8,344.66 (4.37%)	*New London, CT*	4,905.33 (3.50%)	*St. Louis, MO*	11,198.49 (5.33%)
San Diego, CA	6,455.05 (3.38%)	*Dallas, TX*	4,647.67 (3.31%)	*Middlesex, MA*	8,597.06 (4.09%)
Cobb, GA	5,876.74 (3.08%)	*St. Louis, MO*	4,514.87 (3.22%)	*New London, CT*	7,493.77 (3.56%)
Essex, MA	5,202.19 (2.72%)	*Pima, AZ*	4,144.90 (2.95%)	*San Diego, CA*	6,507.77 (3.10%)
St. Louis City, MO	4,485.53 (2.35%)	*King, WA*	4,046.07 (2.88%)	*Pima, AZ*	5,164.25 (2.46%)
Fairfield, CT	4,294.81 (2.25%)	*Fairfield, CT*	3,634.04 (2.59%)	*Fairfield, CT*	5,036.96 (2.40%)
Pima, AZ	4,102.80 (2.15%)	*Cobb, GA*	3,617.65 (2.58%)	*Orange, FL*	4,775.39 (2.27%)
New London, CT	3,805.09 (1.99%)	*Orange, FL*	3,588.32 (2.56%)	*Jackson, MS*	4,416.34 (2.10%)
Maricopa, AZ	3,481.61 (1.82%)	*Los Angeles, CA*	3,558.78 (2.54%)	*King, WA*	4,362.53 (2.08%)
Santa Clara, CA	3,379.90 (1.77%)	*Yolo, CA*	2,758.28 (1.97%)	*Hartford, CT*	4,302.59 (2.05%)
Dallas, TX	3,285.91 (1.72%)	*Orange, CA*	2,580.47 (1.84%)	*Los Angeles, CA*	3,814.94 (1.81%)
Hartford, CT	3,197.50 (1.67%)	*Fairfax, VA*	2,465.61 (1.76%)	*Newport News City, VA*	3,606.09 (1.72%)
Newport News City, VA	3,047.92 (1.60%)	*Santa Clara, CA*	2,427.01 (1.73%)	*Essex, MA*	3,195.11 (1.52%)
Potter, TX	3,031.11 (1.59%)	*Potter, TX*	2,352.76 (1.68%)	*Maricopa, AZ*	3,008.79 (1.43%)
Orange, FL	2,861.36 (1.50%)	*Maricopa, AZ*	2,325.57 (1.66%)	*Cobb, GA*	2,665.60 (1.27%)
King, WA	2,490.13 (1.30%)	*Essex, MA*	2,045.70 (1.46%)	*Winnebago, WI*	2,617.74 (1.25%)
Macomb, MI	2,338.64 (1.22%)	*Newport News City, VA*	2,043.83 (1.46%)	*Orange, CA*	2,609.49 (1.24%)
Middlesex, MA	2,240.90 (1.17%)	*Middlesex, MA*	2,039.87 (1.45%)	*Marion, IN*	2,512.10 (1.19%)

Notes: The table displays the top 20 counties ranked by manufacturing (NAICS 2-digit) DoD procurement expenditures for the years 2010, 2015, and 2020. It includes the corresponding expenditures in millions of dollars, as well as each county's share of the total expenditure (shown in parentheses). All values have been adjusted for inflation using the national defence government expenditure implicit price deflator and are expressed in December 2021 dollars.

Figure B1: County-level manufacturing DoD procurement expenditure (2010)



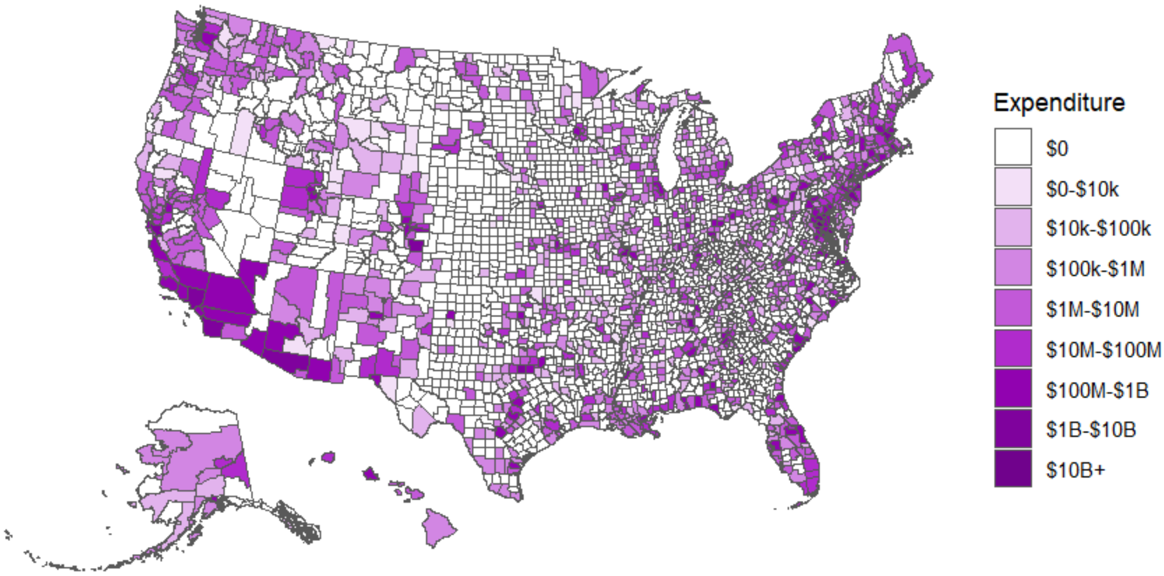
Notes: *The map illustrates the geographical distribution of county-level DoD procurement expenditure for the manufacturing sector (NAICS 2-digit) in 2010. All values have been adjusted for inflation using the national defence government expenditure implicit price deflator and are presented in December 2021 dollars.*

Table B2: Top 20 counties by PSTS DoD procurement expenditure for 2010, 2015, and 2020

2010		2015		2020	
County	Total expenditure (millions of dollars) and share of total	County	Total expenditure (millions of dollars) and share of total	County	Total expenditure (millions of dollars) and share of total
Fairfax, VA	12,815.77 (12.83%)	*Fairfax, VA*	9,803.72 (13.10%)	*Fairfax, VA*	10,678.04 (11.18%)
San Diego, CA	4,569.27 (4.57%)	*Madison, AL*	5,042.27 (6.74%)	*Madison, AL*	5,227.44 (5.48%)
Middlesex, MA	4,352.31 (4.36%)	*District of Columbia, DC*	3,342.01 (4.46%)	*District of Columbia, DC*	4,152.26 (4.35%)
Madison, AL	4,044.48 (4.05%)	*Middlesex, MA*	2,818.28 (3.76%)	*Middlesex, MA*	3,271.86 (3.43%)
Arlington, VA	3,816.73 (3.82%)	*Arlington, VA*	2,427.59 (3.24%)	*Los Angeles, CA*	3,160.63 (3.31%)
District of Columbia, DC	3,634.56 (3.64%)	*San Diego, CA*	2,395.26 (3.20%)	*San Diego, CA*	3,035.86 (3.18%)
Santa Clara, CA	3,099.17 (3.10%)	*Los Angeles, CA*	2,254.23 (3.01%)	*Arlington, VA*	2,832.06 (2.97%)
Monmouth, NJ	2,964.84 (2.97%)	*Allegheny, PA*	2,032.02 (2.71%)	*New London, CT*	2,269.75 (2.38%)
Los Angeles, CA	2,744.09 (2.75%)	*Montgomery, MD*	1,457.16 (1.95%)	*Anne Arundel, MD*	2,160.86 (2.26%)
St. Louis, MO	2,026.76 (2.03%)	*St. Mary's, MD*	1,364.06 (1.82%)	*Santa Clara, CA*	1,954.72 (2.05%)
St. Mary's, MD	1,896.52 (1.90%)	*Orange, FL*	1,093.72 (1.46%)	*El Paso, CO*	1,905.73 (2.00%)
El Paso, CO	1,849.65 (1.85%)	*El Paso, CO*	1,086.87 (1.45%)	*St. Mary's, MD*	1,895.62 (1.99%)
Pima, AZ	1,514.86 (1.52%)	*New London, CT*	1,060.71 (1.42%)	*Allegheny, PA*	1,652.59 (1.73%)
Tarrant, TX	1,481.33 (1.48%)	*Anne Arundel, MD*	1,059.81 (1.42%)	*Pima, AZ*	1,581.04 (1.66%)
Charleston, SC	1,468.46 (1.47%)	*Burlington, NJ*	973.82 (1.30%)	*Montgomery, MD*	1,521.43 (1.59%)
Prince William, VA	1,369.38 (1.37%)	*Prince William, VA*	951.29 (1.27%)	*Howard, MD*	1,456.10 (1.53%)
Howard, MD	1,295.30 (1.30%)	*Howard, MD*	946.83 (1.26%)	*Bexar, TX*	1,441.84 (1.51%)
Burlington, NJ	1,202.29 (1.20%)	*Charleston, SC*	941.37 (1.26%)	*Orange, FL*	1,409.15 (1.48%)
Montgomery, MD	1,052.54 (1.05%)	*Santa Clara, CA*	923.78 (1.23%)	*Norfolk City, VA*	1,306.35 (1.37%)
Allegheny, PA	1,047.33 (1.05%)	*Jackson, MS*	910.41 (1.22%)	*Arapahoe, CO*	1,283.55 (1.34%)

Notes: The table displays the top 20 counties ranked by PSTS (NAICS 2-digit) DoD procurement expenditures for the years 2010, 2015, and 2020. It includes the corresponding expenditures in millions of dollars, as well as each county's share of the total expenditure (shown in parentheses). All values have been adjusted for inflation using the national defence government expenditure implicit price deflator and are expressed in December 2021 dollars.

Figure B2: County-level PSTS DoD procurement expenditure (2010)



Notes: *The map illustrates the geographical distribution of county-level DoD procurement expenditure for the professional, scientific, and technical services sector (NAICS 2-digit) in 2010. All values have been adjusted for inflation using the national defence government expenditure implicit price deflator and are presented in December 2021 dollars.*

Appendix C

R functions for procurement data manipulation

This Appendix describes a series of I developed to compile procurement data from the Department of Defence, sourced from `www.usaspending.gov`. It begins by presenting functions for data aggregation at the national and state levels, followed by functions for county and city levels. The final section includes functions that analyse the data at the company level. The raw data, covering the years 2001 to 2021, is available in the "Procurement data" folder. The codes for the functions are available at `https://github.com/Etexcs/Procurement-data-codes.git`.

Notice that all aggregated expenditure values returned by the functions below are adjusted using the implicit price deflator for government expenditure in national defence from the BEA. However, there are certain cases explicitly mentioned where nominal values are being returned instead.

List of functions

National and State level functions

n_contracts

Description: This function calculates the total number of contracts for a specified year, providing totals for the entire year or on a month-by-month basis. The output can be at either the country level or the state level.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `frequency` – Specifies the frequency of the output data, either as monthly ("m") or yearly ("y").
- `region` – Determines if the returned output is at the country ("country") or state ("state") level.

Example of usage:

```
n_contracts(2007, "y", "state")
```

n_contracts_value

Description: This function counts the total number of contracts in a specified year that has its value above a pre-specified amount. It can return the total number over the year or month by month. The output can be at country or state levels.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `frequency` – Specifies the frequency of the output data, either as monthly ("m") or yearly ("y").
- `region` – Determines if the returned output is at the country ("country") or state ("state") level.

- `value` - Determines the cut-off value.

Example of usage:

```
n_contracts_value(2004, "y", "country", 10000)
```

---

total_value

Description: This function returns the total expenditure in a specified year. It can return the total value over the year or month by month. The output can be at country or state levels.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `frequency` – Specifies the frequency of the output data, either as monthly ("m") or yearly ("y").
- `region` – Determines if the returned output is at the country ("country") or state ("state") level.

Example of usage:

```
total_value(2014, "y", "state")
```

---

total_nominal_value

Description: This function returns the total expenditure in a specified year. It can return the total value over the year or month by month. The values returned are in current dollars and hence not adjusted by inflation. The output can by at country level or state by state.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `frequency` – Specifies the frequency of the output data, either as monthly ("m") or yearly ("y").
- `region` – Determines if the returned output is at the country ("country") or state ("state") level.

Example of usage:

```
total_nominal_value(2002, "m", "country")
```

NAICS_values

Description: This function returns the total expenditure in a specified year, discriminated by the NAICS 2-digit classification. It can return the total value over the year or month by month. The output can be at the country or state levels.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `frequency` – Specifies the frequency of the output data, either as monthly ("m") or yearly ("y").
- `region` – Determines if the returned output is at the country ("country") or state ("state") level.

Example of usage:

```
NAICS_values(2008, "m", "state")
```

NAICS2to3

Description: This function takes a 2-digit NAICS code and return the 3-digit expenditure disaggregation. It can return the total value over the year or month by month. The output can by at country level or for a specified state.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `frequency` – Specifies the frequency of the output data, either as monthly ("m") or yearly ("y").
- `region` – Determines if the returned output is at the country ("country") or state ("state") level.
- `state` – Determines the state to which the data will be returned. Use the abbreviation of the state name. Set as NULL if region = "country".
- `Naics2dig` – The 2-digit NAICS category to be disaggregated. More than one can be used at the same time using the combine (c) function.

Example of usage:

```
NAICS2to3(2009, "m", "country", NULL, 23)
NAICS2to3(2008, "m", "state", "NY", c(48,49))
```

NAICS6dig

Description: This function disaggregates the expenditure of a NAICS 2-digit sector into more granular classifications, moving to a higher number of digits. The function takes a 2-digit sector and the desired number of digits and returns the disaggregated expenditure. The output is at the country level.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `code` – The NAICS 2-digit code of the sector to be disaggregated.
- `digits` – A number from 3 to 6, indicating in how many digits to be used in the disaggregation.
- `cut` – Determines the number of subcategories to be shown. They are ordered by the total expenditure value. Use "ALL" to return all the subcategories.

Example of usage:

```
NAICS6dig(2005, 54, 6, "ALL")
```

NAICS6dig_st

Description: This function disaggregates the expenditure of a NAICS 2-digit sector into more granular classifications, moving to a higher number of digits. The function takes a 2-digit sector and the desired number of digits and returns the disaggregated expenditure. The output is for a pre-determined state.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `code` – The NAICS 2-digit code of the sector to be disaggregated.
- `digits` – A number from 3 to 6, indicating in how many digits to be used in the disaggregation.
- `cut` – Determines the number of subcategories to be shown. They are ordered by the total expenditure value. Use "ALL" to return all the subcategories.

- `state` – Determines the state to which the data will be returned. Use the abbreviation of the state name.

Example of usage:

```
NAICS6dig_st(2012,33,6,20,"OK")
```

PSC_sum

Description: This function returns the total expenditure for products and services classified by the PSC code. It also distinguishes expenditure between R&D and non R&D. Notice all the R&D expenditure are a subcategory of services. It can return the total value over the year or month by month. The output can be at country or state levels.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `frequency` – Specifies the frequency of the output data, either as monthly ("m") or yearly ("y").
- `region` – Determines if the returned output is at the country ("country") or state ("state") level.

Example of usage:

```
PSC_sum(2004, "m", "state")
```

PSC_classf

Description: This function returns the top expenditure classifications for the country or for a given state according to the PSC classification. It is necessary to specify one of the two major categories: "product" or "service". The function will return the total expenditure for all the subcategories within the major category chosen. The returned value is the total over the year. The output can be at the country level or for a specified state.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `region` – Determines if the returned output is at the country ("country") or state ("state") level.
- `state` – Determines the state to which the data will be returned. Use the abbreviation of the state name. Set as NULL if region = "country".
- `type` – Either "product" or "service", determining the major category to be disaggregated.

Example of usage:

```
PSC_classf(2007, "state", "NJ", "product")
PSC_classf(2020, "country", NULL, "service")
```

County-level data functions

total_counties_state

Description: This function returns the total expenditure for all the counties in a specified state in a specified year. It is possible to choose more than one state simultaneously. The function can return the yearly or monthly expenditure.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.

- `frequency` – Specifies the frequency of the output data, either as monthly ("m") or yearly ("y").

- `state` – Determines the state to which the data will be returned. Multiple states can be used simultaneously by using the combine (c) function.

Example of usage:

```
total_counties_state(2008,"y","TN")
total_counties_state(2005,"y",c("OK","TN"))
```

---

total_fips

Description: This function returns the total expenditure for a given county, which is specified by its FIPS code. It is possible to choose more than one county simultaneously. The function can return the yearly or monthly expenditure.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.

- `frequency` – Specifies the frequency of the output data, either as monthly ("m") or yearly ("y").

- `fips` – Numerical argument, determining the FIPS code of the county which the data should be returned. Multiple counties can be used simultaneously by using the combine (c) function.

Example of usage:

```
total_fips(2014,"m",06037)

total_fips(2008,"m",c(06037,40013))
```

NAICS_by_fips

Description: This function takes a 2-digit NAICS code and returns the annual expenditure for each county in the sector specified by that code.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `Naics2dig` – The 2-digit NAICS category to be disaggregated. More than one can be used at the same time using the combine (c) function.

Example of usage:

```
NAICS_by_fips(2010, 42)

NAICS_by_fips(2010, c(31,32,33))
```

NAICS_fips

Description: This function takes a FIPS code for a County and returns the expenditure by 2-digit NAICS classification.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `frequency` – Specifies the frequency of the output data, either as monthly ("m") or yearly ("y").

- `fips` – Numerical argument, determining the FIPS code of the county which the data should be returned. Multiple counties can be used simultaneously by using the combine (c) function.

Example of usage:

```
NAICS_fips(2008, "m", 06037)
```

NAICS2to3_fips

Description: This function takes a 2-digit NAICS code and returns the 3-digit expenditure disaggregation for a given county, which is specified by its FIPS code.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `frequency` – Specifies the frequency of the output data, either as monthly ("m") or yearly ("y").
- `fips` – Numerical argument, determining the FIPS code of the county which the data should be returned. Multiple counties can be used simultaneously by using the combine (c) function.
- `Naics2dig` – The 2-digit NAICS category to be disaggregated. More than one can be used at the same time using the combine (c) function.

Example of usage:

```
NAICS2to3_fips(2006, "y", 6115, 31)
NAICS2to3_fips(8, "y", c(6101,6115), c(42,31))
```

NAICS6dig_fips

Description: This function disaggregates the expenditure of a NAICS 2-digit sector into more granular classifications, moving to a higher number of digits. The function takes a 2-digit sector and the desired number of digits and returns the disaggregated expenditure. The output is for a pre-determined county, selected using its FIPS code.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `code` – The NAICS 2-digit code of the sector to be disaggregated.
- `digits` – A number from 3 to 6, indicating in how many digits to be used in the disaggregation.
- `cut` – Determines the number of subcategories to be shown. They are ordered by the total expenditure value. Use "ALL" to return all the subcategories.
- `fips` – Numerical argument, determining the FIPS code of the county which the data should be returned.

Example of usage:

```
NAICS6dig_fips(12, "ALL", 6, 20, 18067)
```

PSC_counties

Description: This function returns the total expenditure for all the counties in a determined state classified by "product" or "services", accordingly to the PSC classification. It also distinguish expenditure between R&D and non R&D. Notice all the R&D expenditure are a subcategory of services. It can return the total value over the year or month by month.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `frequency` – Specifies the frequency of the output data, either as monthly ("m") or yearly ("y").
- `state` – Determines the state to which the data will be returned. Multiple states can be used simultaneously by using the combine (c) function. The returned output is for each county in the specified states.

Example of usage:

```
PSC_counties(year, frequency, state)
```

PSC_sum_fips

Description: This function returns the total expenditure for products and services classified by the PSC code for a given county. It also distinguish expenditure between R&D and non R&D. Notice all the R&D expenditure are a subcategory of services. It can return the total value over the year or month by month.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `frequency` – Specifies the frequency of the output data, either as monthly ("m") or yearly ("y").
- `fips` – Numerical argument, determining the FIPS code of the county which the data should be returned.

Example of usage:

```
PSC_sum_fips(year, frequency, fips)
```

PSC_fips

Description: This functions returns the top expenditure classifications in a given county according to the PSC classification in one of its two major categories: "product" or "service". The function will return the total expenditure for the subcategories within the major chosen.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `fips` – Numerical argument, determining the FIPS code of the county which the data should be returned.
- `type` – Determines the major category, being either "service" or "product".

Example of usage:

```
PSC_fips(year, fips, type)
```

City-level data functions

---

top_cities

Description: This function returns the cities ranked by expenditure and their respective values in a determined year.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.

Example of usage:

```
top_cities(2014)
```

---

top_cities_state

Description: This function returns cities from a specified state, ranked by expenditure and their respective values for a given year.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `state` – Determines the state to which the data will be returned. Use the abbreviation of the state name.

Example of usage:

```
top_cities_state(2007, "OK")
```

---

top_cities_fips

Description: This function retrieves cities from a specified county, identified by its FIPS code, ranked by expenditure and their respective values in a specific year.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.

- `fips` – Numerical argument, determining the FIPS code of the county which the data should be returned.

Example of usage:

```
top_cities_fips(2010, 06037)
```

---

total_city

Description: This function retrieves the total expenditure for a given city in a determined year. The retrieved data can be at either the monthly or yearly frequency.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `frequency` – Specifies the frequency of the output data, either as monthly ("m") or yearly ("y").
- `city` – Name of the city for which the data will be retrieved. City name needs to be in high case followed by the abbreviation of its state.

Example of usage:

```
total_city(2008, "m", "LONG BEACH, CA")
```

NAICS_city

Description: This function returns the expenditure by 2-digit NAICS classification for the determined city and year.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `city` – Name of the city for which the data will be retrieved. City name needs to be in high case followed by the abbreviation of its state.

Example of usage:

```
NAICS_city(2008, "LONG BEACH, CA")
```

NAICS2to3_city

Description: This function takes a 2-digit NAICS code and returns the 3-digit expenditure disaggregation for a given city.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `frequency` – Specifies the frequency of the output data, either as monthly ("m") or yearly ("y").
- `city` – Name of the city for which the data will be retrieved. City name needs to be in high case followed by the abbreviation of its state.
- `naics2dig` – The 2-digit NAICS category to be disaggregated. More than one can be used at the same time using the combine (c) function.

Example of usage:

```
NAICS2to3_city(2018,"y", "LOS ANGELES, CA", c(31,32,33))
```

PSC_sum_city

Description: This function returns the total expenditure for products and services classified by the PSC code for a given city. It also distinguish expenditure between R&D and non R&D. Notice all the R&D expenditure are a subcategory of services. It can return the total value over the year or month by month.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `frequency` – Specifies the frequency of the output data, either as monthly ("m") or yearly ("y").
- `city` – Name of the city for which the data will be retrieved. City name needs to be in high case followed by the abbreviation of its state.

Example of usage:

```
PSC_sum_city(2011, "m", "ORLANDO, FL")
```

PSC_city

Description: This functions returns the top expenditure classifications in a given city according to the PSC classification in one of its two major categories: "product" or "service". The function will return the total expenditure for the subcategories within the major chosen.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.

- `city` – Name of the city for which the data will be retrieved. City name needs to be in high case followed by the abbreviation of its state.

- `type` – Determines the major category, being either "service" or "product".

Example of usage:

```
PSC_city(2005, "LOS ANGELES, CA", "product")
```

Company-level functions

---

total_supplier

Description: This function returns the total expenditure for a company, identified by its Unique Entity Identifier (UEI code), in a given year. The expenditure is also discriminated by the city in which the company operates.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.

- `uei` – The UEI code to identify the business.

- `frequency` – Specifies the frequency of the output data, either as monthly ("m") or yearly ("y").

Example of usage:

```
total_supplier(2009, "W9EQK2KSXTW4", "y")
```

top_companies

Description: This function returns the top companies in expenditure and their respective values for the specified year.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `cut` – Determines how many companies are shown.

Example of usage:

```
top_companies(2007,30)
```

top_companies_st

Description: This function returns the top companies in expenditure and their respective values in a determined state and year.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `state` – Determines the state to which the data will be returned. Use the abbreviation of the state name.
- `cut` – Determines how many companies are shown.

Example of usage:

```
top_companies_st(2020, "OK", 50)
```

top_companies_fips

Description: This function returns the top companies in expenditure and their respective values in a determined county and year.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `fips` – Numerical argument, determining the FIPS code of the county which the data should be returned.
- `cut` – Determines how many companies are shown.

Example of usage:

```
top_companies_fips(2008,26005,5)
```

top_companies_city

Description: This function returns the top companies in expenditure and their respective values in a determined city and year.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `city` – Name of the city for which the data will be retrieved. City name needs to be in high case followed by the abbreviation of its state.
- `cut` – Determines how many companies are shown.

Example of usage:

```
top_companies_city(2018, "FORT WORTH, TX", 50)
```

top_companies_NAICS

Description: This function returns the top companies for a given NAICS code. It accepts codes with 2 or 3 digits.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `naics` – The 2 or 3 digit NAICS code for the desired sector.
- `cut` – Determines how many companies are shown.

Example of usage:

```
top_companies_NAICS(2018, 115, 50)
```

UEItoNAICS

Description: This function takes the UEI code of a business and a year and returns the expenditure of that company ranked by the 2-digit NAICS classification.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `uei` – The UEI code to identify the business.

Example of usage:

```
UEItoNAICS(2010,"FLK3MUWVKFG3")
```

PSC_companies

Description: This function takes a PSC code and a specified year to return the top companies and their respective expenditures for that year and category.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `psc` – The PSC for the desired category.
- `cut` – Determines how many companies are shown.

Example of usage:

```
PSC_companies(2005,34,20)
```

---

location_company

Description: This function pinpoints the location of the expenditure related to a given company. The company can be identified by its name or its UEI code using the "type" argument. The company can be identified either by its name or its UEI code using the "type" argument. It returns the expenditures broken down by each location (city and county levels) of the company for the specified year.

Arguments:

- `year` – Numeric argument, from 2001 to 2021, indicating the year of the data to be compiled.
- `type` – Either "UEI" or "Name", depending on what is being used to identify the company.
- `identifier` – The UEI code or company's name, depending on what was chosen in the "type" argument.

Example of usage:

```
location_company(2008,"UEI","LKTZWE6LNEB6")
```

Appendix D

LMA estimations

Table D1: Estimation results (spillover effects)
Employment, labour force, and population

Outcome Variable	**Employment**		**Labour Force**		**Population**	
Specification	**(4.3)**	**(4.4)**	**(4.3)**	**(4.4)**	**(4.3)**	**(4.4)**
Expenditure	0.127*** (0.003)	0.074*** (0.003)	0.112*** (0.003)	0.066*** (0.002)	0.109*** (0.003)	0.065*** (0.002)
D^{high} * Expenditure		0.341*** (0.016)		0.307*** (0.015)		0.289*** (0.014)
Expenditure LMA	0.003 (0.005)	-0.027*** (0.005)	0.018*** (0.005)	-0.019*** (0.004)	0.016*** (0.005)	-0.016*** (0.004)
$D^{high,LMA}$ * Expenditure LMA		0.086*** (0.023)		0.103*** (0.022)		0.082*** (0.020)
D^{high}		-4.921*** (0.267)		-4.526*** (0.251)		-4.212*** (0.232)
$D^{high,LMA}$		-1.651*** (0.478)		-1.943*** (0.467)		-1.518*** (0.425)
Counties	3087		3122		3130	
Years	16		15		16	
Observations	49392		46830		50080	

Notes: The table displays the estimation results for equations (4.3) and (4.4), focusing on employment, the labour force, and population as the outcome variables. The first row of the table specifies the outcome variable, while the second row presents the estimated equation. All variables in the regression have been transformed using an inverse hyperbolic sine transformation. Labour Market Areas (LMA), as per the 1990 census data, were used to define local labour markets. The variables denoted by LMA represent the values for the LMA the county belongs to, excluding itself. Each regression includes a varying number of observations based on the outcome variable, as noted in the last row. All regressions account for both time and state fixed effects, and standard errors are clustered at the county level. The symbols "***", "**", and "*" indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table D2: Estimation results
Outcome variable : Employment per capita

Specification	**(4.3)**	**(4.4)**
PC Expenditure	0.012*** (0.001)	0.011*** (0.001)
$Dhigh$ * PC Expenditure		0.000 (0.003)
PC Expenditure LMA	-0.005*** (0.001)	-0.005*** (0.001)
$D^{high,LMA}$ * PC Expenditure LMA		0.005 (0.004)
D^{high}		0.010 (0.014)
$D^{high,LMA}$		-0.042 (0.026)
Counties	3087	
Years	16	
Observations	49392	

Notes: The table displays the estimation results for equations (4.1) to (4.4), using average wages as the outcome variable. The estimated equation is indicated in the first row of the table. Notice that equations (4.3) and (4.4) were estimated using Labour Market Areas (LMA), as per the 1990 census data, to determine the local labour market. The variables denoted by LMA represent the values for the LMA the county belongs to, excluding itself. All variables in the regression have been transformed using an inverse hyperbolic sine transformation. Each regression includes a varying number of observations based on the outcome variable, as noted in the last row. All regressions account for both time and state fixed effects, and standard errors are clustered at the county level.
The symbols "***", "**", and "*" indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table D3: Estimation results
Outcome variable: average weekly wages

Specification	**(4.3)**	**(4.4)**
Expenditure	0.008*** (0.000)	0.003*** (0.000)
D^{high} * Expenditure		0.043*** (0.003)
Expenditure LMA	0.003*** (0.001)	-0.003*** (0.001)
$D^{high,LMA}$ * Expenditure LMA		0.021*** (0.004)
D^{high}		-0.659*** (0.044)
$D^{high,LMA}$		-0.397*** (0.087)
Counties	3087	
Years	16	
Observations	49392	

Notes: The table displays the estimation results for equations (4.1) to (4.4), using average wages as the outcome variable. The estimated equation is indicated in the first row of the table. Notice that equations (4.3) and (4.4) were estimated using Labour Market Areas (LMA), as per the 1990 census data, to determine the local labour market. The variables denoted by LMA represent the values for the LMA the county belongs to, excluding itself. All variables in the regression have been transformed using an inverse hyperbolic sine transformation. Each regression includes a varying number of observations based on the outcome variable, as noted in the last row. All regressions account for both time and state fixed effects, and standard errors are clustered at the county level.
The symbols "***", "**", and "*" indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Appendix E

Estimations including county FE

Table E1: Estimation results
Employment, labour force, and population

Outcome Variable	Employment		Labour Force		Population	
Specification	**(4.3)**	**(4.4)**	**(4.3)**	**(4.4)**	**(4.3)**	**(4.4)**
Expenditure	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
D^{high} * Expenditure		0.005*** (0.001)		0.000 (0.002)		0.004*** (0.001)
D^{high}		-0.071*** (0.023)		0.009 (0.026)		-0.053*** (0.016)
Counties	3087		3122		3130	
Years	16		15		16	
Observations	49392		46830		50080	

Notes: The table presents the estimation results for equations (4.1) and (4.2), with employment, labour force, and population as the outcome variables. The first row indicates the outcome variable, while the second row the estimated equation. All variables in the regression have been transformed using an inverse hyperbolic sine transformation. Each regression includes a varying number of observations based on the outcome variable, as noted in the last row. All regressions account for a year, state, and county fixed effect, and standard errors are clustered at the county level. The symbols "***", "**", and "*" indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table E2: Estimation CMZ spillovers (county FE)
Employment, labour force, and population

Outcome Variable	**Employment**		**Labour Force**		**Population**	
Specification	**(4.3)**	**(4.4)**	**(4.3)**	**(4.4)**	**(4.3)**	**(4.4)**
Expenditure	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
D^{high} * Expenditure		0.005*** (0.002)		0.000 (0.001)		0.004*** (0.001)
Expenditure CMZ	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.001** (0.001)	0.000 (0.000)	0.000 (0.000)
$D^{high,CMZ}$ * Expenditure CMZ		-0.002 (0.003)		-0.018*** (0.003)		0.000 (0.001)
D^{high}		-0.007*** (0.002)		0.008 (0.026)		-0.052*** (0.016)
$D^{high,CMZ}$		0.050 (0.058)		0.366*** (0.056)		-0.006 (0.029)
Counties	3080		3122		3080	
Years	16		15		16	
Observations	49280		46830		49280	

Notes: The table displays the estimation results for equations (4.3) and (4.4), focusing on employment, the labour force, and population as the outcome variables. The first row of the table specifies the outcome variable, while the second row presents the estimated equation. All variables in the regression have been transformed using an inverse hyperbolic sine transformation. Commuting zones (CMZ), as per the 2020 census data, were used to define local labour markets. The variables denoted by CMZ represent the values for the CMZ the county belongs to, excluding itself. Each regression includes a varying number of observations based on the outcome variable, as noted in the last row. All regressions account for a year, state, and county fixed effect, and standard errors are clustered at the county level.
The symbols "***", "**", and "*" indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table E3: Estimation results
Outcome variable: employment per capita (county FE)

Specification	**(4.1)**	**(4.2)**	**(4.3)**	**(4.4)**
PC Expenditure	0.001 (0.001)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
D^{high} * PC Expenditure		0.001 (0.002)		0.000 (0.000)
PC Expenditure CMZ			0.000 (0.000)	0.000 (0.000)
$D^{high,CMZ}$ * PC Expenditure CMZ				0.001 (0.001)
D^{high}		-0.005 (0.008)		-0.001 (0.004)
$D^{high,CMZ}$				-0.004 (0.005)
Counties	3087		3080	
Years	16		16	
Observations	49392		49280	

Notes: The table displays the estimation results for equations (4.1) to (4.4), using employment per capita as the outcome variable. The estimated equation is indicated in the first row of the table. Notice that equations (4.3) and (4.4) were estimated using Commuting Zones (CMZ), as per the 2020 census data, to determine the local labour market. The variables denoted by CMZ represent the values for the CMZ the county belongs to, excluding itself. All variables in the regression have been transformed using an inverse hyperbolic sine transformation. Each regression includes a varying number of observations based on the outcome variable, as noted in the last row. All regressions account for a year, state, and county fixed effect, and standard errors are clustered at the county level.
The symbols "***", "**", and "*" indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table E4: Estimation results
Outcome variable: average weekly wages (County FE)

Specification	**(4.1)**	**(4.2)**	**(4.3)**	**(4.4)**
Expenditure	0.001** (0.000)	0.001*** (0.000)	0.001** (0.000)	0.001*** (0.000)
D^{high} * Expenditure		-0.003*** (0.001)		-0.003*** (0.001)
Expenditure CMZ			0.000 (0.000)	0.000 (0.000)
$D^{high,CMZ}$ * Expenditure CMZ				-0.007*** (0.002)
D^{high}		0.047*** (0.014)		0.042*** (0.014)
$D^{high,CMZ}$				0.139*** (0.030)
Counties	3087		3080	
Years	16		16	
Observations	49392		49280	

Notes: The table displays the estimation results for equations (4.1) to (4.4), using average wages as the outcome variable. The estimated equation is indicated in the first row of the table. Notice that equations (4.3) and (4.4) were estimated using Commuting Zones (CMZ), as per the 2020 census data, to determine the local labour market. The variables denoted by CMZ represent the values for the CMZ the county belongs to, excluding itself. All variables in the regression have been transformed using an inverse hyperbolic sine transformation. Each regression includes a varying number of observations based on the outcome variable, as noted in the last row. All regressions account for a year, state, and county fixed effect, and standard errors are clustered at the county level.
The symbols "***", "**", and "*" indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Appendix F

NAICS 2-digit mapping into primary sectors

Table F1: NAICS 2-digit mapping into primary sectors

Code	NAICS 2-digit category	Primary Sector
11	*Agriculture, forestry, fishing, and hunting*	*Natural resources*
21	*Mining*	*Natural resources*
22	*Utilities*	*Services*
23	*Comstruction*	*Construction*
31,32,33	*Manufacturing*	*Manufacturing*
42	*Wholesale trade*	*Services*
44,45	*Retail trade*	*Services*
48,49	*Transportation and warehousing*	*Services*
51	*Information*	*Services*
52	*Finance and insurance*	*Services*
53	*Real estate rental and leasing*	*Services*
54	*Professional, scientific and technical services*	*Services*
55	*Management of companies and enterprises*	*Services*
56	*Administrative and support and waste management and remediation services*	*Services*
61	*Educational services*	*Services*
62	*Health care and social assistance*	*Services*
71	*Arts, entertainment, and recreation*	*Services*
72	*Accommodation and food services*	*Services*
81	*Other services, except government*	*Services*
92	*Public administration*	*Services*

Notes: The table shows how each of the 21 NAICS 2-digit categories was allocated into the four primary sectors: Construction, Manufacturing, Natural Resources, and Services.