Impacts of E-commerce on Warehousing and Distribution in California

February 2025

A Research Report from the Pacific Southwest Region University Transportation Center

Genevieve Giuliano, University of Southern California

Seula Lee, University of Southern California

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. PSR-23-01 TO 079	**2. Government Accession No.** N/A	**3. Recipient's Catalog No.** N/A
4. Title and Subtitle Impacts of E-commerce on warehousing and distribution in California		**5. Report Date** 04/03/2025 **6. Performing Organization Code** N/A
7. Author(s) Genevieve Giuliano, 0000-0002-9257-8269 Seula Lee, 0000-0002-1002-0553		**8. Performing Organization Report No.** PSR-23-01 TO 079
9. Performing Organization Name and Address METRANS Transportation Center University of Southern California University Park Campus, RGL 216 Los Angeles, CA 90089-0626		**10. Work Unit No.** N/A **11. Contract or Grant No.** USDOT Grant 69A3551747109 Caltrans Agreement 65A0674, TO-79
12. Sponsoring Agency Name and Address U.S. Department of Transportation Office of the Assistant Secretary for Research and Technology 1200 New Jersey Avenue, SE, Washington, DC 20590		**13. Type of Report and Period Covered** Final report (1/1/2024-12/31/2024) **14. Sponsoring Agency Code** USDOT OST-R
15. Supplementary Notes https://doi.org/10.25554/ygj6-9775		
16. Abstract The purpose of this research is to document and analyze trends in location patterns of warehousing and distribution (WD) activity in California over the past decade, and to explore the relationship between these trends and the growth of e-commerce. This research builds on a previous study of WD trends in California 2003-2013 and extends to 2022.[1] The research has two parts. Part 1 is a descriptive analysis of WD trends, Part 2 estimates models to explain these trends. There was an approximate doubling of WD establishments over the period, but the overall spatial distribution of activity was markedly stable. There is no evidence of decentralization; growth took place throughout the state's metro areas. We estimate both cross section and time series models, finding that local market attributes consistently explain WD location. Transport access plays a less significant role. We conclude that continued growth even in high density core areas is consistent with the rapid growth in e-commerce that took place over the same period.		
17. Key Words Warehousing and distribution, e-commerce,	**18. Distribution Statement** No restrictions.	

19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 75	**22. Price** N/A

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

[1] Giuliano, G. and S. Kang (2017) Spatial Dynamics of Warehousing and Distribution in California, Final Report, METRANS UTC 15-27. Available at https://www.metrans.org/assets/research/15-27%20%2865A0533%20TO%20016%29%20Final%20Report%20pdf.pdf.

Pacific Southwest Region UTC
University Transportation Center

Contents

List of Tables

List of Figures

About the Pacific Southwest Region University Transportation Center

The Pacific Southwest Region University Transportation Center (UTC) is the Region 9 University Transportation Center funded under the US Department of Transportation's University Transportation Centers Program. Established in 2016, the Pacific Southwest Region UTC (PSR) is led by the University of Southern California and includes seven partners: Long Beach State University; University of California, Davis; University of California, Irvine; University of California, Los Angeles; University of Hawaii; Northern Arizona University; Pima Community College.

The Pacific Southwest Region UTC conducts an integrated, multidisciplinary program of research, education and technology transfer aimed at *improving the mobility of people and goods throughout the region.* Our program is organized around four themes: 1) technology to address transportation problems and improve mobility; 2) improving mobility for vulnerable populations; 3) Improving resilience and protecting the environment; and 4) managing mobility in high growth areas.

U.S. Department of Transportation (USDOT) Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

California Department of Transportation (CALTRANS) Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the United States Department of Transportation's University Transportation Centers program, in the interest of information exchange. The U.S. Government and the State of California assumes no liability for the contents or use thereof. Nor does the content necessarily reflect the official views or policies of the U.S. Government and the State of California. This report does not constitute a standard, specification, or regulation. This report does not constitute an endorsement by the California Department of Transportation (Caltrans) of any product described herein.

Disclosure

Principal Investigator Genevieve Giuliano and Seula Lee conducted this research titled, "Impacts of E-commerce on Warehousing and Distribution in California, at the Department of Urban Planning and Spatial Analysis, University of Southern California. The research took place from March 15, 2024 to December 31, 2024 and was funded by a grant from the California Department of Transportation in the amount of $99,999. The research was conducted as part of the Pacific Southwest Region University Transportation Center research program.

Acknowledgements

We acknowledge funding from the California Department of Transportation. We thank Caltrans Task Order Manager Connor Campbell and his team for excellent project management and assistance with Caltrans data sources. We thank Prof. Sanggyun Kang for providing data files from a previous project. The assistance of graduate research assistants Caleb Pak and Sharvari Rajwaday in compiling literature, collecting data, and creating database files is greatly appreciated. All errors and omissions are the responsibility of the authors.

Abstract

The purpose of this research is to document and analyze trends in location patterns of warehousing and distribution (WD) activity in California over the past decade, and to explore the relationship between these trends and the growth of e-commerce. This research builds on a previous study of WD trends in California 2003-2013 and extends to 2022.[2] The research has two parts. Part 1 is a descriptive analysis of WD trends, Part 2 estimates models to explain these trends. There was an approximate doubling of WD establishments over the period, but the overall spatial distribution of activity was markedly stable. There is no evidence of decentralization; growth took place throughout the state's metro areas. We estimate both cross section and time series models, finding that local market attributes consistently explain WD location. Transport access plays a less significant role. We conclude that continued growth even in high density core areas is consistent with the rapid growth in e-commerce that took place over the same period.

[2] Giuliano, G. and S. Kang (2017) Spatial Dynamics of Warehousing and Distribution in California, Final Report, METRANS UTC 15-27. Available at https://www.metrans.org/assets/research/15-27%20%2865A0533%20TO%20016%29%20Final%20Report%20pdf.pdf.

Impacts of E-commerce on Warehousing and Distribution in California

Executive Summary

The purpose of this research is to document and analyze trends in location patterns of warehousing and distribution (W&D) activity in California over the past decade, and to explore the relationship between these trends and the growth of e-commerce. E-commerce has grown rapidly over the past two decades, with an average annual US growth rate of about 11% until the COVID pandemic, and increasing over 50% during the pandemic. After mid 2020 e-commerce sales continued to increase, but at a lower average growth rate. As of 2024, the US e-commerce market share is 15% and the global share is 20%. With its emphasis on short delivery times and an ever-expanding array of products, new forms of supply chains have emerged which may have significant impacts on W&D demand and location choice. On the one hand, scale economies in warehousing leads to demand for peripheral locations where land is cheaper and more available. On the other hand, short delivery times require access to the population. Therefore, impacts on the spatial distribution of W&D are unclear.

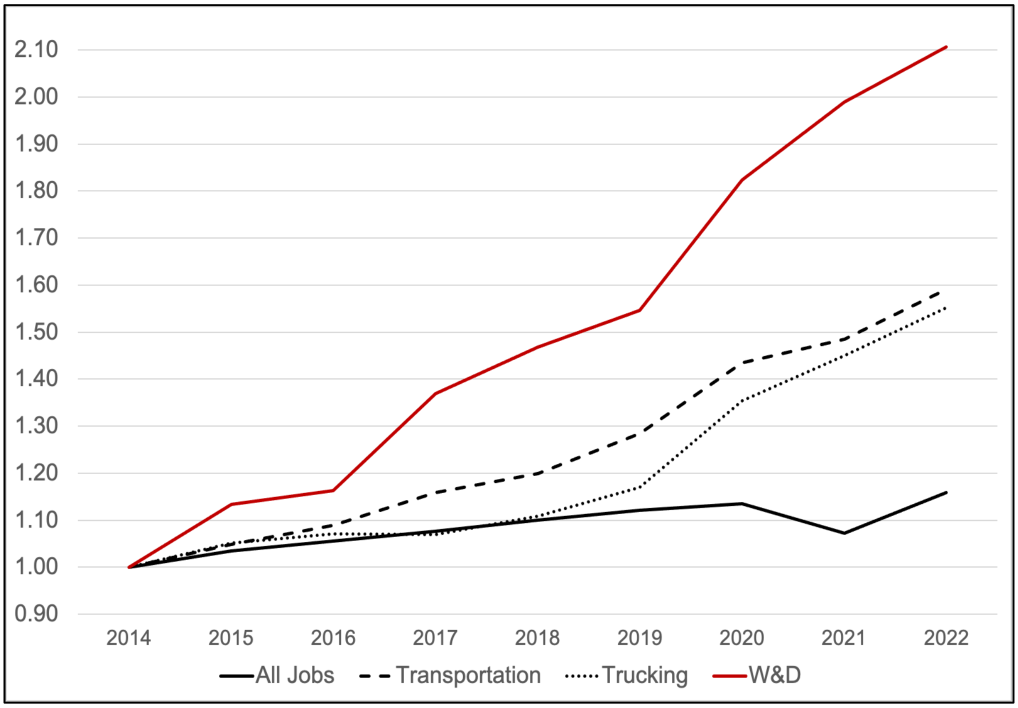
This research builds on a previous study of W&D trends in California 2003-2013 and extends to 2022.[3] The research has two parts. Part 1 is a descriptive analysis of WD trends, Part 2 estimates models to explain these trends. Our primary data sources are the County Business Patterns (CBP) and Zip Code Business Patterns (ZBP) data compiled by the US Census. These sources provide annual data on establishments and employment (CBP only) at up to the 6-digit NAICS (North American Industry Classification System) level. We use 3-digit data to examine W&D growth and spatial dynamics from 2014 to 2022. The zip code level data is converted to Zip Code Tabulation Areas (ZCTAs) created by the US Census to map zip code data.

Results

Growth of W&D has far outpaced that of the general economy or the broader transportation sector as illustrated in Figure E-1. The increases over the period are 15% for all employment, 60% for the 2-digit transportation sector, and 111% for the 3-digit W&D sector. Given the approximate doubling of the sector, the spatial distribution could have changed dramatically as well.

[3] Giuliano, G. and S. Kang (2017) Spatial Dynamics of Warehousing and Distribution in California, Final Report, METRANS UTC 15-27. Available at https://www.metrans.org/assets/research/15-27%20%2865A0533%20TO%20016%29%20Final%20Report%20pdf.pdf.

Figure E 1 Trends in Relative Job Growth, Entire Economy and Subsectors, 2014-2022



We examined spatial trends at varying levels of geography. We used US Census defined MSAs (Metropolitan Statistical Area) and MiSA (Micropolitan Statistical Area) to categorize California's metropolitan and rural areas by size. We use four levels as described in Table E 1. The level 1 MSAs experienced the greatest increase in W&D employment and establishments, generating a slight increase in share for the largest MSAs and slight decreases in share for the other categories.

Table E 1 Study Area Categorization

Level	Definition	N
1	MSAs with population greater than 2 million	6 MSAs
2	MSAs with population 250,000 to 2 million	13 MSAs
3	MSAs with population less than 250,000	7 MSAs
4	MiSAs and rural counties	8 MISAs, 13 counties

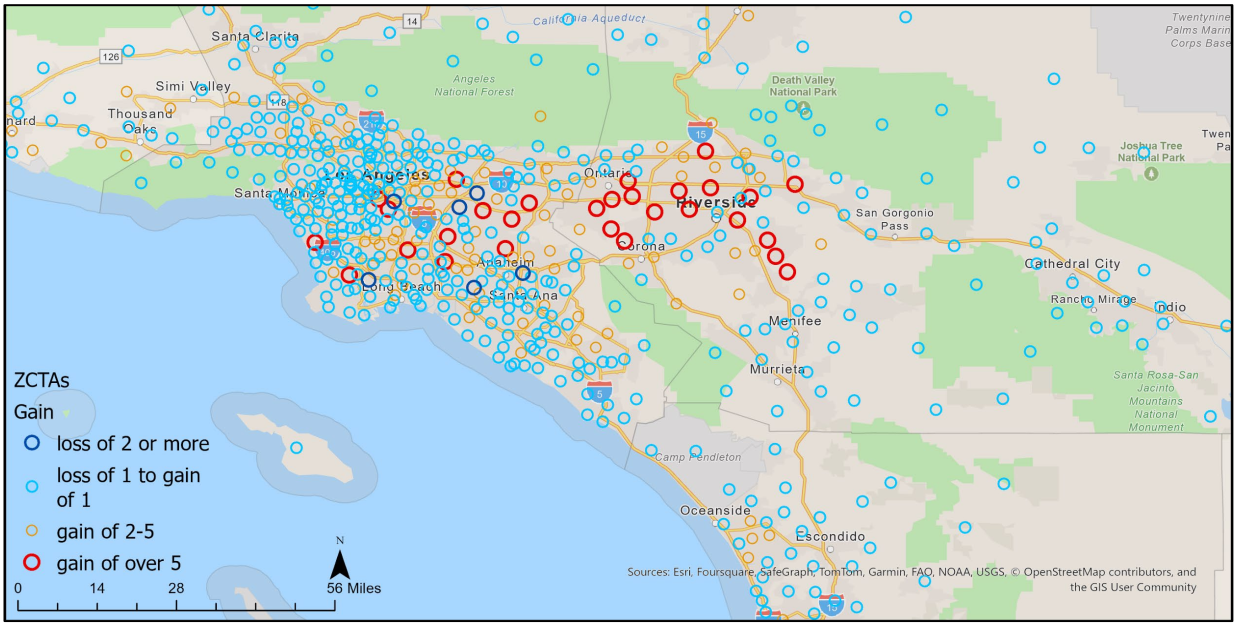
We explored spatial trends using the ZCTA establishment data (employment at the 3-digit level is not available at the zip code level). Our findings are of remarkable stability: the growth of the sector has largely followed the existing spatial pattern. Figure E 2 provides an example for the greater Los Angeles region. It can be seen that gains are distributed throughout -- from the coastline to the Inland Empire. Stability is also demonstrated by calculating the weighted average distance of all W&Ds to the CBD (Central Business District), identified as the ZCTA with the highest employment density in the MSA. None of the differences from 2014 to 2022 were statistically significant.

We conducted a statistical analysis to examine factors associated with W&D location. Based on the literature we develop measures for local market, regional market and transport access factors. We estimate three models: binomial for whether or not a ZCTA has at least one W&D; negative binomial and zero-inflated negative binomial for the count of W&D establishments in a ZCTA. We estimate cross section models for 2014, 2019, and 2022; time series models for 2014-2019, 2019-2022, and 2014 – 2022. All models include spatial lags to control for the spatial correlation in the dependent variable. We use 2019 to separate pre- and post- COVID effects.

Results across the cross-section models were largely consistent. Coefficients for local market variables are generally significant and have the expected positive sign, meaning that W&Ds are more likely located in areas with more labor force access. Our regional market measure is never significant in the binary and NB (Negative Binomial) models. Results for transport access measures are mixed. Coefficients for airport access are generally significant and of the expected sign. Positive signs for distance to seaports and intermodal facilities are explained by their geographic locations. In several cases, the coefficient of highway access variable is positive, which is counterintuitive. We empirically observe that nearly 80% of all W&Ds are within one mile of a highway and this proportion has actually increased slightly over time. It is possible that the correlation between the independent variables affects our results.

Our findings for the time series models are similar to those for the cross-section models. Within each model form there is consistency in the results, with the possible exception of a slight difference for the 2019-2022 comparison. The same pattern of significance for local market and regional market factors is observed. Coefficients for access to airports are consistently significant. Overall, transport access measures do not explain much of the variation in W&D location patterns.

Figure E 2 W&D Establishment Gains and Losses, Los Angeles Region, 2014-2022



Conclusions

The spatial organization of warehousing is remarkably stable: the industry sector more than doubled over the time period of our study and the additional activity simply intensified the existing pattern. This spatial stability is illustrated in our region level maps of W&Ds, the demonstrated absence of change in location with respect to the CBD, and the general lack of significance of transport access variables in our statistical analysis. It appears that the decentralization or spillovers observed in the previous 2003 – 2013 study have played out – those peripheral areas are now part of the spatial pattern, but few new distant locations have emerged. This process of “infill” growth is consistent with e-commerce related demands for access to the population and short delivery times. It is also consistent with the increasing velocity of supply chains more generally. Overall spatial stability is also explained by the concentration of population and jobs in a few very large metropolitan areas, the role of the largest metropolitan areas in the national and international economies, and path dependence driven by infrastructure investments and historical growth patterns.

Chapter 1: Introduction

The California economy is one of the largest in the world. With an estimated equivalent gross domestic product of $3.9 trillion in 2023, it ranks 5th among the world's economies.[4] California remains the top state for manufacturing by value of total output.[5] California seaports, airports and land connections together make California the nation's top international trade gateway, with approximately $628 billion in trade in 2023.[5] California's large and dynamic economy, together with its role as the nation's major international trade gateway, generates large volumes of freight flows and an active warehousing and distribution sector.

California's role in domestic production and international trade is layered on the demands of a population of nearly 40 million whose consumption patterns continue to change as e-commerce grows. E-commerce is increasing around the world. In the US, e-commerce is increasing about 11% annually, much greater than the rate of total retail sales, (3-4%), and the market share is now 15%.[6] The global market share is estimated to be 20% and valued at about $4.2 trillion USD in 2024.[7] The emergence of online shopping has transformed where and how goods are produced, distributed, and sold, and how consumers make shopping as well as shopping travel decisions.

E-commerce is changing rapidly. The variety of goods available continues to grow, and many new products have emerged, such as customized beauty products and subscription deliveries of frequently used products. Speed of delivery is also increasing. Large online retailers offer 'instant deliveries' (within two hours) in some cities, and one-day delivery is now routine in many metropolitan areas. More recent changes include prepared food deliveries from industrial kitchens and the emergence of individual deliveries by cars (Uber Eats) or bicycle (Grubhub).

The growth of e-commerce has impacts on both transportation and urban form. First, freight flows become increasingly fragmented as more retail products are delivered to individuals rather than retail stores, delivery times shrink, and more products get delivered as individual shipments. Fragmentation increases vehicle miles traveled (VMT). Increased VMT in turn generates more congestion, air pollution, and energy consumption. Second, to fulfill short delivery times, retailers must be as close to customers as possible. In-city warehouse and distribution space is therefore in high demand, but land prices and other constraints limit the

[4] Source: https://www.ppic.org/publication/californias-economy/#:~:text=California's%20economy%20ranks%20fifth%20internationally,of%20all%20of%20these%20countries.

[5] Bureau of Transportation Statistics, https://www.bts.gov/content/top-us-foreign-trade-freight-gateways-value-shipments-current-billions.

[6] US Census, https://www.census.gov/retail/data.html

[7] Sources: Statista, https://www.statista.com/topics/871/online-shopping/; International Trade Administration, US Department of Commerce, https://www.trade.gov/ecommerce-sales-size-forecast.

size and number of facilities. The supply chain has responded with more complex distribution networks: large facilities in less urbanized areas serving as hubs for in-city distribution.[8]

California therefore has two forces for increased goods movement: growing international trade and growing demand for consumer deliveries. It is therefore timely to examine how these two forces may be restructuring supply chains and changing patterns of WD location, which in turn affect goods movement on the transportation system.

Many of the factors that affect the location of W&Ds are those that generally affect all profit-maximizing firms. For W&Ds, the trade-offs are between land costs, transport costs, inventory costs, labor and other inputs. All else equal, firms will select the combination of these factors that minimizes total costs or maximizes profits. Land price plays a major role; firms may trade off transport costs for cheaper land. Location shifts may occur as relative costs change over time. For example, population and economic growth increase land rents as demand for land intensifies.

All else equal, we would expect W&D – a land intensive activity – to shift away from areas with increasing rents and seek new locations in less developed areas. Transport costs also play a significant role. Access to major trade nodes – major highways, port, airport and intermodal terminals – is essential to fulfilling global freight demands. There are three factors unique to W&Ds that may lead to decentralization of W&D location. First, the industry itself is changing rapidly. Scale economies, generated by information systems and automation, are increasing demand for very large-scale facilities (McKinnon, 2009), which intensifies demand for low land prices and large parcels. Second, structural shifts in the supply chain affect W&Ds. Examples include incorporating secondary processes in distribution, increasing the velocity of supply chains, and omni-channel retail distribution systems (McKinnon, 2009; Napolitano, 2013). Third, the environmental impacts associated with W&Ds affect more people in densely developed areas. Local opposition may act as a push factor for relocation of W&D activity to less developed areas. However, access to customers to serve quick delivery markets is a countervailing force; some form of fulfillment center must be located close enough to accomplish one day or less deliveries. With the growth of e-commerce, it is therefore possible that spatial location trends are changing.

Trends in W&Ds are of interest for the following reasons. First, W&Ds are major truck traffic generators. If location patterns are shifting over time, their associated truck travel demand will also shift, affecting the highway system. Understanding how and why these shifts are taking place is essential for metropolitan and statewide planning. Second, factors affecting W&Ds suggest fewer but larger scale operations, located further from population centers, and agglomeration economics suggest the development of large warehouse clusters. More

[8] Rodrigue, J-P (2020) The distribution network of Amazon and the footprint of freight digitization, Journal of Transport Geography, 88, 102825, https://doi.org/10.1016/j.jtrangeo.2020.102825..

concentration implies greater localized impacts. On the other hand, e-commerce demands suggest at least some activity within dense population centers.

The contrasting forces of scale and access to population likely lead to increased VMT. The focus on velocity and highly flexible supply chains may affect mode choice in favor of trucking. Rail transport is slower, less flexible, and reliant on large shipment size, but at the same time more energy efficient. Within the truck mode, these trends may lead to use of smaller trucks and more frequent trips as deliveries become increasingly customized and dispersed. Given California's greenhouse gas (GHG) emission reduction goals, it is important to understand the underlying dynamics of truck demand so that appropriate policies can be designed to effectively manage demand.

The purpose of this research is to examine W&D location trends in California from 2014 to 2022. It builds on a previous analysis of trends from 2003 to 2013 (Giuliano and Kang, 2017). The previous study found that overall W&D activity is distributed approximately with the population and employment, with some evidence that W&D activity was moving away from major metropolitan areas and significant decentralization observed for the San Francisco and Los Angeles metropolitan regions. The study concludes that absent major external shocks, W&Ds will remain concentrated in the largest metro areas, and those in less populated areas will continue to cluster around high access nodes of the highway network.

This research follows the same approach. We conduct a descriptive analysis of location trends, then estimate models to explain these trends. The remainder of this report is organized as follows. Chapter two presents a review of the recent literature on W&D location. Chapter three presents the descriptive analysis of spatial and temporal trends at various levels of geography. Chapter four presents results from model estimations to test hypotheses regarding factors associated with the observed trends. The final chapter presents conclusions and discusses the policy implications of the results.

Chapter 2: Literature Review

2.1 Introduction

Location of warehousing and distribution facilities (WDs) may be best understood in the context of traditional location theory (e.g. Isard, 1956; Losch, 1954; Moses, 1958): firms locate to minimize costs of inputs of production and transport of products to markets. Access to inputs, land prices, labor force access, access to markets, taxes and regulations, as well as agglomeration economies, are some of the key factors in industrial location choice.

Traditionally, WDs located near input sources – ports, rail terminals, manufacturing clusters – because of the relatively high cost of moving bulky materials. WDs served to divide large product shipments and deliver in smaller scale lots to retailers. Over the past decades much has changed. Continued growth and decentralization of metropolitan areas changed the employment and population distributions and created higher (but flatter) land price gradients. These changes pushed land intensive activities to the periphery. Within the WD industry itself, scale economies associated with automation, as well as the restructuring of supply chains associated with the emergence of e-commerce, have led to the demand for ever larger facilities, in turn increasing demand for large land parcels.

The literature on WD location focuses on two spatial trends: decentralization (often termed logistics sprawl) and clustering. A second theme is externalities and environmental justice. This review is organized around these themes. We provide a brief summary of work before 2015 that was previously presented in Giuliano and Kang (2017), and a more comprehensive discussion of research since 2015.

2.2 Summary of literature before 2015

The growth of WDs led to several studies in the early 2000s. Much of it focused on the development of large facilities in suburban or exurban locations, signifying a fundamental change in the warehousing and distribution process in part driven by the emergence of e-commerce. Many studies of spatial trends were conducted, and results were mixed. A study of WD patterns in The Netherlands indicated increased concentration (van den Heuval et al, 2013); studies of Los Angeles, Atlanta, and the UK indicated decentralization (Dablanc and Ross, 2012; Dablanc, et al., 2014; Allen, Browne and Cherrett, 2012). One national study found a general trend of decentralization (Cidell, 2010). Differences in results can be attributed at least in part to how decentralization is measured. Giuliano, Kang and Yuan (2015) conducted a study of WD location in the four largest metro areas in California – Los Angeles, San Francisco, Sacramento and San Diego – from 2003 - 2013. They used four different measures: average distance to the CBD, average distance to all employment, Gini coefficient, and share of WDs located in the first (highest) quartile of employment density. Significant decentralization was found only in the case of Los Angeles.

The California work was expanded by Giuliano and Kang (2017). Using zip code business patterns data from 2003 to 2013, they conducted a two-part analysis. The first is a descriptive analysis. They find that 1) the W&D industry grew much faster than other parts of the economy, 2) W&D activity is distributed approximately with the population and total employment, 3) there is some evidence of W&D activity moving away from the major metro areas to nearby smaller metro areas; 4) significant decentralization is observed for Los Angeles and San Francisco. The second part estimates models of WD location choice. Labor force access is consistently significant over the period, as is access to highways and intermodal facilities. In contrast, measures of inter-industry linkages decline in significance over the period. The authors conclude that while WD location patterns are quite stable overall, there is some indication of spillover effects into smaller metro areas adjacent to Los Angeles and San Francisco.

2.3 Research from 2015

2.3.1 Decentralization

The question of logistics sprawl continues to be a major topic, and results have become more consistent. WD decentralization and its impacts have been documented in large metropolitan areas around the world, including Tokyo (Sakai, Kawamura and Hyodo, 2015), Wuhan (Yuan and Zhu, 2019), Brussels (Strale, 2020), Gothenburg (Heitz et al, 2018), Zurich (Todesco and Weidmann, 2016), Toronto (Woudsma, Jacobicek, and Dablanc, 2016), and Chicago (Dubie et al, 2020). The one exception is Phoenix, but the large geographic units used in the study may explain the results (Dubie et al, 2020). These results suggest a strong structural dynamic in the logistics chain that is global in nature.

Southern California has been the focus of several studies. Jaller and Pineda (2017) use a combination of Zipcode Business Pattern (ZBP) and the Commodity Flow Survey (CFS) microdata to examine WD patterns from 1998 to 2014 in the Los Angeles region. They calculate barycenters (weighted geometric centers) and Gini coefficients and find a general trend of decentralization through 2007. Their results are consistent with Giuliano and Kang (2018). Jaller, Qian and Zhang (2020, 2022) use industry building sales data to examine location trends in Southern California from 1998 to 2018. They compare building size and price over time and space, finding evidence of re-centralization in recent years with more sales and higher prices for more central locations. The trend is consistent with logistic shifts related to e-commerce.

Kang (2020b) conducts one of the few national level studies of WD decentralization. He uses ZBP data from 2003 to 2016 for the 64 largest US metropolitan areas. The question is whether WDs are decentralizing more than other employment or population – whether WD decentralization is simply part of the larger trend in urban spatial structure. Results show that WDs have decentralized somewhat more than other related industries and population in the period before 2010. Much of the shift is attributable to the largest facilities in the largest metro areas – places where land price would have the most influence. Kang (2020c) uses employment density as a proxy for land price and confirms that decentralization is most evident for the largest WD facilities in the densest metro areas.

2.3.2 Recent trends and e-commerce

A few recent papers have focused on trends related more specifically to e-commerce. Dablanc (2019) describes recent logistics trends such as instant deliveries, non-motorized delivery services, and food preparation kitchens and their potential impacts on urban logistics. Hesse (2020) provides an overview of trends in supply chains and distribution and their impacts. The rise of e-commerce has restructured supply and distribution chains, leading to the emergence of clusters of large facilities at greenfield sites and local fulfillment centers. With about half of the online market, Amazon is the major player in the design of these distribution networks.

Rodrigue (2020) describes a new “freight landscape” of e-commerce built on functional specialization. The distribution system begins with cross-dock facilities at major import nodes, from which goods are shipped to e-fulfillment centers. Goods are assembled to orders; orders go to sortation centers and then to parcel delivery stations for route distribution. Each of these facilities has different location imperatives. A case study of the US Amazon network shows that the large-scale cross dock and e-fulfillment centers locate at the periphery of large metro areas, while smaller sortation and delivery stations locate in the core.

2.3.3 Factors related to location choice

A major research question is what is driving these spatial patterns? Referring back to the industry location literature, which factors are more important, and has their importance changed over time? Onstein et al (2018) conduct a comprehensive literature review on warehouse location as part of a general review of logistics distribution structures. Significant factors include access to highways and intermodal facilities, labor and land availability, taxes and local regulatory environment.

Table 1 below summarizes findings from several studies. Results are strikingly consistent despite different types of data, time frames, and study location. Access to highways and intermodal facilities, inter-industry linkages or agglomeration economies, land prices and access to population are significant factors. The Paris study shows the importance of strong land use controls. WDs seek locations as close as possible to population centers, but high land prices (and possibly restrictive zoning) push them to lower density locations with good access to highways and intermodal facilities.

Table 1 Summary: Explanatory factors for WD location

Author/date	Method	Data	Results
Ginerich and Maoh, 2019	Logit models	489 WDs in Toronto area, 2015	Presence of industry, access to highways and airports, land price
Holl and Mariotti, 2018	Various forms of logit models	New WD locations in Spain, 2002-2007	Access to highways and airports, access to major metro areas, inter-industry linkages not significant when metro area variable included
Jaller and Pineda, 2017	Spatial regression models	Commodity Flow Survey microdata, 1998-2014, Los Angeles region	Presence of manufacturing and retail, access to highways and intermodal facilities, agglomeration economies, household income (-)
Kang, 2020a	Various forms of logit models	WDs built before 1980 and after 2000, Los Angeles region	Access to local markets, labor force and intermodal facilities for pre-1980; land price, access to airports and intermodal facilities for post-2000 facilities
Sakai, Beziat and Heitz, 2020	Various logit models	826 logistics facilities in Paris region, 2003-2013	Agglomeration economies, land use policy, land price, access to highways
Yang et al., 2022	Zero-inflated negative binomial models	New logistics facility locations in Shanghai, 2005-2015	Access to intermodal facilities, industry linkages, population density, (access to highway ramp not significant)

2.3.4 Environmental impacts

One of the potential impacts of WD decentralization is increased truck VMT. It is argued that WD decentralization moves distribution further from customers and hence generates more VMT. However, if customers (e.g. population) are also decentralizing, this may not be the case. Kang (2020a) finds that WDs have decentralized somewhat more than population or employment, but the trend is limited to the largest metro areas and before 2010.

A major challenge in addressing the question of VMT is the lack of data. The Tokyo study (Sakai, Kawamura and Hyodo, 2015) had access to shipment data and was able to estimate differences in shipping distances, showing that longer shipping distances are associated with more distant locations. Rivera-Royero, Jaller and Kim (2021) use truck flow data from weight-in-motion stations in Southern California from 2003-2015 to examine truck flow patterns. They find that medium and light duty truck traffic increased more than heavy duty truck traffic, suggesting more local deliveries associated with the rise in e-commerce. Because the truck flows could not be directly linked to WD location trends, we cannot conclude that the increased VMT is associated with WD location trends.

A second consideration is the environmental justice aspects of WD location. There is a growing literature on this topic. See Yuan (2018b) for a recent review. The growth of large facilities and distribution networks have resulted in greater demand for lower priced land sufficiently near population centers. Lower income, minority communities are more likely to be in or near such locations.

Most of the empirical studies have focused on Southern California. DeSousa, Ballare, and Niemeier (2022) examine environmental risks and traffic impacts of warehousing in Southern California. Using a data set of 3321 large warehouses (over 100,000 sq. ft.), they map the number of warehouses in each census tract. They collect data on PM 2.5 exposure, noise, vehicle collisions, and traffic density to test whether these risks are associated with presence of warehouses. They estimate spatial regression models and find a weak association between warehouses and exposure.

Jaller, Qian and Zhang (2020, 2022) use ZBP data for the 5 largest MPOs in California to compare location trends and local population characteristics. They find that WD location is correlated with areas scoring as high pollution burden as measured by California's CalEnviroScreen.

Yuan (2019a) uses WD leasing data to conduct a cross-sectional analysis of WD location and population characteristics for the four largest metropolitan areas in California. Spatial regression models revealed a consistently significant relationship between WD location and medium income minority neighborhoods. In other work he shows how local planning practice influences WD location; cities that perceive the need for economic development may be more willing to accept new WD development (Yuan, 2019b). In Yuan (2021), WD location models are estimated to test whether WD location in the Los Angeles region is more likely in low income, minority neighborhoods; results show that WD location is more likely near medium income minority neighborhoods.

A major question in WD location is the direction of causality. Several studies have shown an association between WD location and minority, low-income neighborhoods. Which came first? Do WD developers seek our minority neighborhoods, or do minority populations seek out lower housing prices, trading off the noise and air pollution? Yuan (2018a) uses simultaneous equations to test causality using the Los Angeles region as a case study. Results show that WDs are locating in minority neighborhoods, likely because of lower land prices and a more permissive local approval process.

A neighborhood level case study was conducted to examine the impacts of a new online grocery delivery warehouse in the South Bronx. Using quasi-experimental design, Shearston et al (2020) compare traffic level, noise, and air quality (black carbon and PM2.5) before and after the start of operation of the warehouse. They find increased truck traffic and noise, and slightly increased air pollution, adding to the burden of this environmental justice community.

2.4 Conclusion

The literature on WD location shows that the trends observed in California are replicated around the world. Structural changes in the logistics sector have driven demand for larger facilities and hence for locations with lower land prices. Decentralization forces are strongest in the largest metro areas, where land prices are high. Access to highways and intermodal facilities allows for more efficient distribution networks including for the last mile. There is little evidence as yet regarding “re-centralization” to facilitate one day or instant deliveries.

Chapter 3: Descriptive analysis

3.1 Introduction

This research builds on the previous analysis (Giuliano and Kang, 2017) and follows a similar methodological approach. We analyze the trends in W&D distribution in California in two parts. In Part I, we describe trends over the last decade – change in overall numbers of W&Ds at multiple geographic levels, change in W&D distribution with respect to general employment and population trends, and change in W&D spatial patterns. In Part II, we assess multiple explanatory factors associated with these trends. Several statistical models test the extent to which the factors explain the cross-sectional distribution and its changes over time. Here we specify the research framework of the first part.

Because the state of California is diverse in terms of its development density, we delineate the region into four levels of geography. The first three levels are metropolitan areas of different population size categories, and the fourth level includes micropolitan and rural counties. Based on this delineation, we describe the distribution and changes in the number of W&Ds at three different geographic scales – the entire state, four metro levels, county, and ZIP Code. Then, we identify areas of growth or decline and compare trends. To evaluate whether W&D spatial trends simply replicate the larger spatial trends of the entire economy, we compare the numbers of W&Ds to the numbers of total establishments and employment. If so, we may conclude that location choice factors are similar, and population and employment growth would be good proxies for predicting future patterns. If not, we are interested in how and why W&D patterns differ, and what implications these may have for truck travel. To examine extent of concentration or de-concentration we calculate location quotients and distance to the central business district (CBD) over time.

3.2 Delineating the study area

As in the previous study, we begin by delineating regions within the state. The state is very diverse; it is home to Los Angeles, the second largest metropolitan area in the US, as well as the metro areas of San Francisco – San Jose, San Diego, and Sacramento. The state also has vast agricultural regions as well as sparsely populated desert and forest regions. The previous study used Consolidated statistical areas (CSAs) as the basis for delineation of regions. These were based on the 2010 US Census. The 2020 Census redefined the boundaries of the San Francisco CSA to San Francisco, San Jose, Sacramento, and all the counties connecting them. For our purposes this CSA is too diverse and geographically too big. We have therefore chosen to use Metropolitan Statistical Areas (MSAs) and Micropolitan Statistical Areas (MiSAs) to segment the urban parts of the state. A MSA consists of one or multiple adjacent counties with at least one urban area with more than 50,000 population; a MiSA consists of one or more counties with one urban area with 10,000-50,000 population. Neighboring counties are combined to form an MSA, if the level of social and economic interactions (quantified by commuting ties) is over the threshold OMB (Office of Management and Budget) designates. Any counties that are not MSAs or MiSAs are rural. We use MSAs, MiSAs and rural areas together with population cutoffs and group counties as follows:

- Level 1: MSA with population greater than 2 million
- Level 2: MSA with population greater than 250,000 and less than 2 million
- Level 3: MSA with population less than 250,000
- Level 4: MiSA or rural county

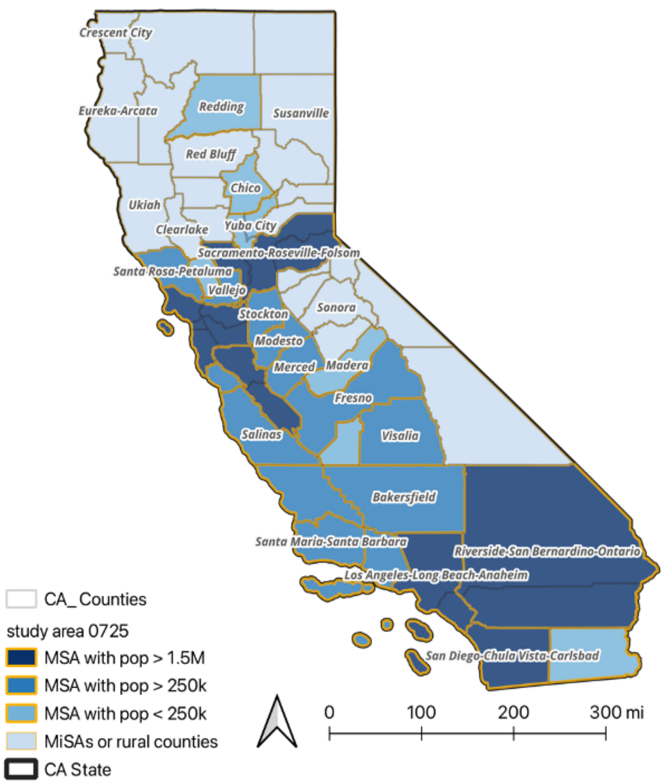
Table 2 lists all MSAs, MiSAs and rural counties by level, and Figure 1 maps their location. The 6 largest MSAs (level 1) account for just over 76% of the state population. Level 2 accounts for 18%, level 3 for 3%, and level 4 for 2%, Figure 1 shows that population is concentrated in the greater Los Angeles and San Francisco areas and along the interior areas between them.

Table 2 Study Area MSAs, MiSAs, and Rural Counties by Level

Level	Full Name	Type	Population in 2020 (thousand)
1 MSAs with population greater than 2M	Los Angeles-Long Beach-Anaheim	MSA	13,201
	San Francisco-Oakland-Berkeley	MSA	4,749
	Riverside-San Bernardino-Ontario	MSA	4,600
	San Diego-Chula Vista-Carlsbad	MSA	3,298
	Sacramento-Roseville-Folsom	MSA	2,397
	San Jose-Sunnyvale-Santa Clara	MSA	2,000
2 MSAs with population between 2M and 250K	Fresno	MSA	1,008
	Bakersfield	MSA	909
	Oxnard-Thousand Oaks-Ventura	MSA	844
	Stockton	MSA	779
	Modesto	MSA	553
	Santa Rosa-Petaluma	MSA	489
	Visalia	MSA	473
	Vallejo	MSA	453
	Santa Maria-Santa Barbara	MSA	448
	Salinas	MSA	439
	San Luis Obispo-Paso Robles	MSA	282
	Merced	MSA	281
	Santa Cruz-Watsonville	MSA	271
3 MSAs with population less than 250k	Chico	MSA	212
	Redding	MSA	182
	Yuba City	MSA	181
	El Centro	MSA	180
	Madera	MSA	156
	Hanford-Corcoran	MSA	152
	Napa	MSA	138
4	Eureka-Arcata	MiSA	136
	Truckee-Grass Valley	MiSA	102

MiSAs and rural counties	Ukiah	MiSA	91
	Clearlake	MiSA	68
	Red Bluff	MiSA	66
	Sonora	MiSA	56
	Susanville	MiSA	32
	Crescent City	MiSA	28
	All rural counties – Siskiyou, Modoc, Trinity, Glenn, Colusa, Plumas, Sierra, Alpine, Amador, Calaveras, Mariposa, Mono, Inyo	Rural counties	279

Figure 1 Selected Geography for Regional Comparisons



3.3 Data

The primary data sources are the US Census' County and ZIP Code Business Patterns (CBP and ZBP) data. The data are based on the Business Register in which records of every known business with an EIN (employer identification number) are maintained. CBP and ZBP provide the number of establishments at the 6-digit industry code level. We use NAICS 493 'Warehousing and Storage' to identify W&D establishments. The Census Bureau defines 'establishments' as "a single physical location at which business is conducted, or services or industrial operations are performed."[9] ZBP is based on USPS ZIP Codes. CBP and ZBP data are reported annually.

3.3.1 Data problems and adjustments

The CBP data gives the total number of employees, total number of establishments, and number of establishments by size category for each county and for each 6-digit NAICS. For our purpose we need data at the 3-digit NAICS level. However, the data are censored when there are small numbers, resulting in a large amount of missing data, especially in counties with low numbers of W&D establishments. The ZBP data gives total number of establishments for each zip code, but no employment data at the 3-digit NAICS level. Therefore, any analysis of 3-digit sector employment at the zip code level requires using some method of imputing employment from the establishment data.

The time period for this analysis is 2014 through 2022. US Census changed its rules regarding censoring of data in 2017. From 2017, any county or zip code with 3 or less establishments has an establishment entry of missing. To minimize the missing data problems, we imputed values where possible by using historical data. For example, if a given county or zip code has an entry of missing in 2017 for the number of establishments but has a valid number in adjacent previous years, we use a 3-year average of those numbers for the missing year. When we imputed numbers for a zip code within a given county, we adjust the imputed number so that the county total would remain consistent for that year. These imputations allowed us to preserve all the establishment data for zip codes. After the data imputation process, the discrepancy between the county-level data and the aggregated zip code data for the entire sample was less than 1% throughout the study period (2014-2022). However, missing data increases as the number of establishments in a zip code decrease. For MiSAs and rural counties, missing data is extensive even prior to 2017. Thus, there is more error associated with the level 4 county imputations.

Our analysis is constrained by these data problems. For our descriptive analysis we use both employment and establishment data at the county or MSA level, but only establishment data at the zip code level. For our statistical analysis, we delete MiSAs and rural counties.

We make an additional adjustment with the ZBP data. While the ZBP data are spatially identified by zip code, zip codes are not officially spatially defined. Rather, zip codes are designed around delivery areas for a given post office. In addition, other national data files such as the American Community Survey do not provide data by zip code. The US Census has created ZIP Code Tabulation Areas (ZCTAs) to map zip code data. Although ZCTAs have 5-digit numbers like zip codes, there is not a one-to-one match between them. Some zip codes are simply post

[9] https://www.census.gov/programs-surveys/susb/about/glossary.html#:~:text=An%20establishment%20is%20a%20single,Establishment%20Births.

office boxes, others represent a single business rather than a geographic location. We convert the zip code data to ZCTAs for spatial analysis.

3.4 General Trends at the State Level

We present descriptive statistics of W&D trends in California in comparison to the entire economy and the transportation sector. Tables 3.2 and 3.3 give annual establishments and employment respectively for the entire economy, the transportation two-digit sector (NAICS 48-49), truck transportation (NAICS 484), and warehousing and storage (NAICS 493).

The 2014-2022 period was one of steady growth for the California economy; employment and establishments grew about 15%. The transportation sector grew much faster. As a result, the transportation share increased by about one percentage point, from 3.3 to 4.6% of jobs and 2.5 to 3.8% of establishments. Trucking and W&D grew even faster. Trucking jobs increased by about 55% and establishments more than doubled; W&D jobs more than doubled and establishments increased by about 43%. These numbers suggest growing numbers of smaller firms in trucking and growing numbers of larger firms in W&D. Both sectors increased their share of the state's economy.

Table 3 Annual Employment and Establishments, Entire State Economy and Transportation Sector

Year	The entire economy		NAICS 48-49 Transportation		NAICS 48-49 share of total economy	
	Jobs	Est.	Jobs	Est.	Jobs	Est.
2014	13,838,702	889,646	463,483	22,057	3.35%	2.48%
2015	14,325,377	908,120	486,149	23,153	3.39%	2.55%
2016	14,600,349	922,477	505,066	23,852	3.46%	2.59%
2017	14,896,625	941,377	536,987	24,840	3.60%	2.64%
2018	15,223,664	954,632	555,804	26,004	3.65%	2.72%
2019	15,516,824	966,224	595,328	27,193	3.84%	2.81%
2020	15,710,859	981,369	664,832	30,300	4.23%	3.09%
2021	14,835,360	998,582	688,325	34,270	4.64%	3.43%
2022	16,032,440	1,023,181	737,237	38,589	4.60%	3.77%
Change	15.85%	15.01%	59.06%	74.95%	37.30%	52.12%

Table 4 Annual Employment and Establishments, Trucking and W&D

Year	NAICS 484 Truck transportation		Share of		NAICS 493 Warehousing and Storage		Share of	
	Jobs	Est.	Jobs	Est.	Jobs	Est.	Jobs	Est.
2014	105,510	9,820	0.76%	1.10%	86,936	1,846	0.63%	0.21%
2015	110,980	10,540	0.77%	1.16%	98,565	1,924	0.69%	0.21%
2016	113,011	11,284	0.77%	1.22%	101,052	1,969	0.69%	0.21%
2017	112,822	11,879	0.76%	1.26%	119,069	2,153	0.80%	0.23%
2018	116,958	12,935	0.77%	1.35%	127,689	2,182	0.84%	0.23%
2019	123,372	13,738	0.80%	1.42%	134,450	2,238	0.87%	0.23%
2020	142,916	15,774	0.91%	1.61%	158,509	2,251	1.01%	0.23%
2021	152,933	19,584	1.03%	1.96%	172,978	2,376	1.17%	0.24%
2022	163,777	23,298	1.02%	2.28%	183,150	2,648	1.14%	0.26%
Change	55.22%	137.25%	33.98%	106.29%	110.67%	43.45%	81.85%	24.72%

Figures 2 and 3 provide indexed graphs of employment growth for the state economy, the transportation sector, trucking, and W&D. The base years are 2014 and 2003 respectively. Figure 2 shows the COVID-related recession for the state economy, but the pandemic years generated rapid growth for the other sectors. The trend is consistent with the shifts in consumer demand and increased online shopping observed during the pandemic. We used the data from the previous Giuliano and Kang (2017) study to generate a longer time series. During 2003-2013 the impact of the Great Recession is evident with all sectors in decline from 2008 through 2010. W&D pulls away around 2012 and continues steep growth through 2022. Over two decades W&D employment has tripled. While the growth in establishments is not nearly as steep, these trends indicate significant impact on the landscape as more and larger W&D facilities are built.

Figure 2 Trends in Relative Job Growth, Entire Economy and Subsectors, 2014-2022

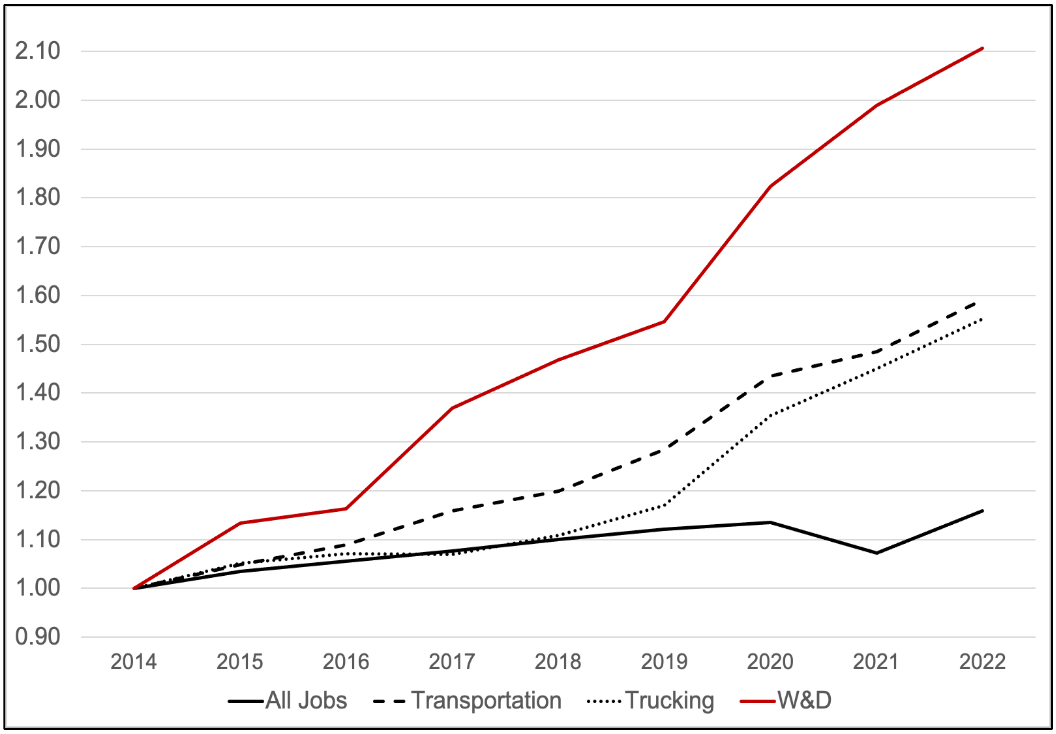
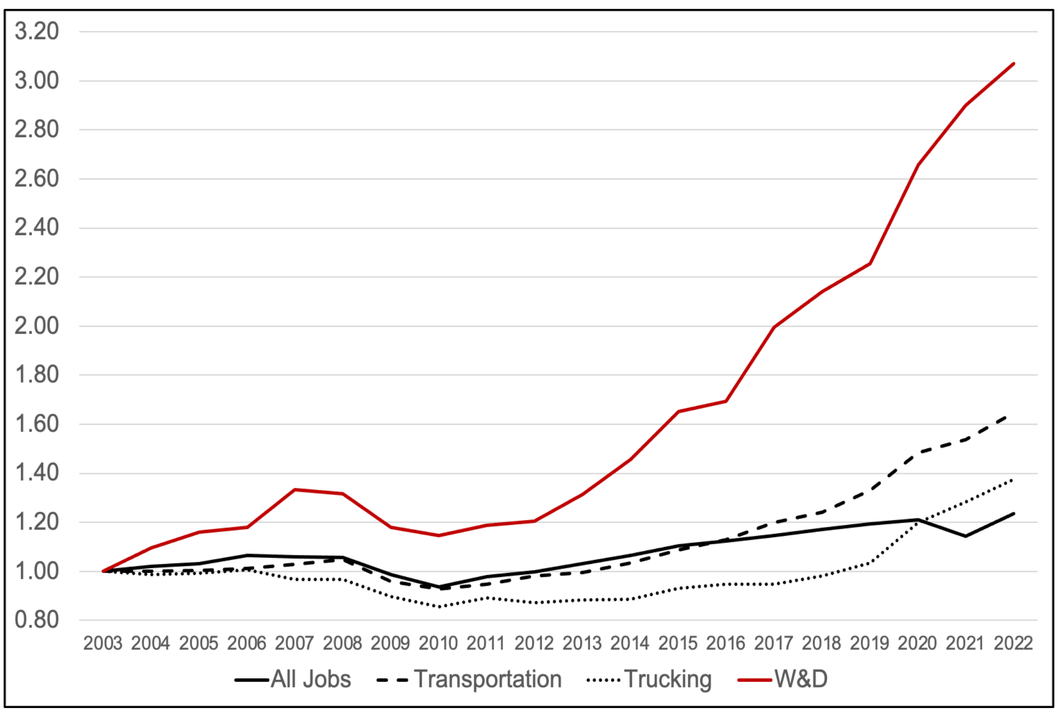


Figure 3 Trends in Relative Job Growth, Entire Economy and Subsectors, 2003-2022



3.5 Trends at the regional level

Tables 5 and 6 give the number of establishments and employment respectively for the entire economy, transportation, and W&D by metropolitan level. The following observations are drawn from the tables. First, the vast share of economic activity – over 80% -- takes place in the state's six largest MSAs, slightly higher than the share of population these MSAs represent. For other metropolitan levels, the economic activity share is slightly lower than the population share. Second, the distribution of economic activity across metropolitan levels is stable over time for total activity and the transportation sector. There is a slight increase in relative share in W&D for the largest MSAs. This is an indication of increased concentration. Finally, establishment and employment trends are quite similar.

Table 5 Total, Transportation, and W&D Establishments by Metropolitan Level, 2014 and 2022

Level	The entire economy			
	2014		2022	
	N	Share	N	Share
1	716,722	80.63%	828,707	81.12%
2	132,129	14.86%	149,627	14.65%
3	21,296	2.40%	23,441	2.29%
4	18,752	2.11%	19,750	1.93%
Total	888,899		1,021,525	
Level	**Transportation**			
	2014		**2022**	
	N	**Share**	**N**	**Share**
1	16,886	76.59%	29,223	75.78%
2	3,914	17.75%	7,451	19.32%
3	778	3.53%	1,318	3.42%
4	470	2.13%	569	1.48%
Total	22,048		38,561	
Level	**W&D**			
	2014		**2022**	
	N	**Share**	**N**	**Share**
1	1,432	77.62%	2,124	79.61%
2	315	17.07%	434	16.27%
3	71	3.85%	82	3.07%
4	27	1.46%	28	1.05%
	1,845		2,668	

Table 6 Total, Transportation, and W&D Employment by Metropolitan Level, 2014 and 2022

Level	The entire economy				Transportation			
	2014		2022		2014		2022	
	N	Share	N	Share	N	Share	N	Share
1	11,239,125	83.36%	12,451,056	83.11%	382,161	82.70%	576,640	81.37%
2	1,792,661	13.30%	2,034,266	13.58%	67,369	14.58%	115,017	16.23%
3	277,338	2.06%	309,922	2.07%	7,707	1.67%	11,269	1.59%
4	173,475	1.29%	186,293	1.24%	4,882	1.06%	5,769	0.81%
Total	13,482,599		14,981,537		462,119		708,695	
Level	W&D				Share of population in 2020			
	2014		2022					
	N	Share	N	Share				
1	66,946	77.96%	145,900	80.41%	77.04%			
2	17,982	20.94%	34,251	18.88%	18.43%			
3	939	1.09%	1,298	0.72%	3.06%			
4	1,443	1.68%	1,347	0.74%	1.48%			
	85,867		181,449					

Table 7 shows changes in establishments and jobs by metropolitan level. Overall, establishments for the total economy and the transportation sector increased more than jobs, but for the W&D sector jobs increased more than establishments, suggesting larger scale facilities. For the total economy, there is no clear trend across metropolitan levels, except that micropolitan and rural areas had substantially lower growth. For the transportation sector, level 2 metropolitan areas grew the most and level 4 grew the least. Level 1 metro areas experienced the greatest growth in W&D, again suggesting continuing concentration.

Table 7 Changes in Establishments and Jobs by Metropolitan Level

Level	All businesses		Transportation		W&D	
	Est.	Jobs	Est.	Jobs	Est.	Jobs
1	15.62%	10.78%	73.06%	50.89%	48.32%	117.94%
2	13.24%	13.48%	90.37%	70.73%	37.71%	90.47%
3	10.07%	11.75%	69.41%	46.22%	15.49%	38.23%
4	5.32%	7.39%	21.06%	18.17%	3.70%	6.65%
Total	14.92%	11.12%	74.90%	53.36%	44.61%	113.31%

Another way to compare W&D dynamics across metropolitan levels is to examine relative concentration: are W&Ds more concentrated than total economic activity across metropolitan size categories? We use the Location Quotient to measure relative concentration. The Location Quotient (LQ) quantifies the spatial concentration of an industry in a region (Miller et al., 1991).

LQ is the ratio of two shares: the share of employment in industry (i) in metro area (j) relative to total employment in metro area (j); and the share of employment in industry (i) in California relative to total California employment. It is calculated as follows:

$$LQ = \frac{\frac{Emp_{i,j}}{Emp_j}}{\frac{EMP_i}{EMP}} \qquad (1)$$

Where,

$Emp_{i,j} = \text{N of employment in industry } i \text{ in metro area } j$

$Emp_j = \text{N of all employment in metro area } j$

$EMP_i = \text{N of employment in industry } i \text{ in California}$

$EMP = \text{N of all employment in California}$

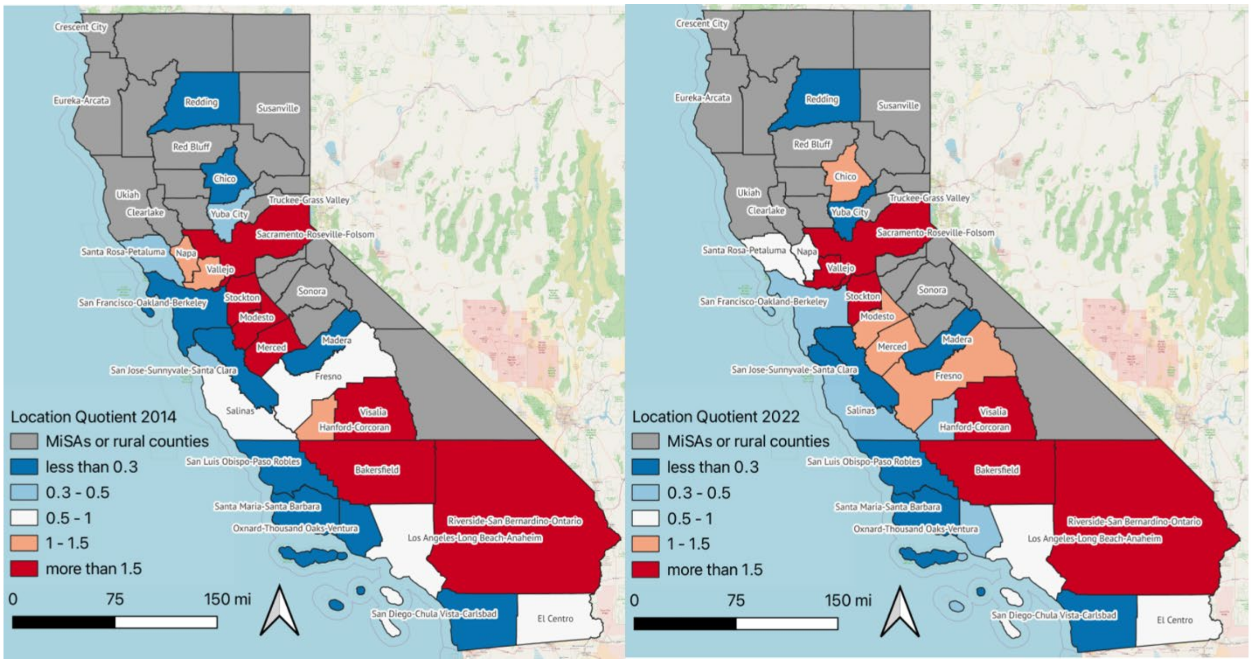
The LQ is interpreted as follows. If LQ is equal to 1, the share of the given industry in a given metropolitan area is the same as its share in the California economy. If LQ is greater than one, the relative share is greater than that of California, and if LQ is less than one the relative share is smaller than that of California. Table 8 gives results. For the transportation sector, the LQ is close to one for levels 1 and 2, but less than 1 for levels 3 and 4. W&D is notably more concentrated in level 2, but not for level 1. We expect level 1 to have an LQ close to one because it accounts for such a large share of California employment. W&D is much less concentrated in level 3, and concentration is declining. Although level 2 has the highest relative concentration of W&D, concentration has declined over the 2014 – 2022 period. W&D for level 4 is not shown due to missing employment data.

Table 8 Location Quotient by Metropolitan Level, 2014 and 2022

Level	Transportation			W&D		
	2014	2022	% change	2014	2022	% change
1	0.993	0.979	-1.40%	0.923	0.955	3.49%
2	1.094	1.195	9.22%	1.555	1.373	-11.70%
3	0.810	0.769	-5.06%	0.525	0.341	-34.93%
4	0.786	0.655	-16.76%	N/A	N/A	N/A
Total	1.00	1.00		1.00	1.00	

Figure 5 shows the LQ for W&D by MSA or county for 2014 and 2022, respectively. The grey areas are the level 4 counties with no W&D employment data. Each map has the same scale: blue shades to white are scores from less than one to one; red shades are scores over 1. The highest concentrations in both years are in the inland counties surrounding Los Angeles and in the Sacramento area. There is also substantial concentration with some mixed changes in the Central Valley. In contrast, W&D facilities are less concentrated in the central coastal counties.

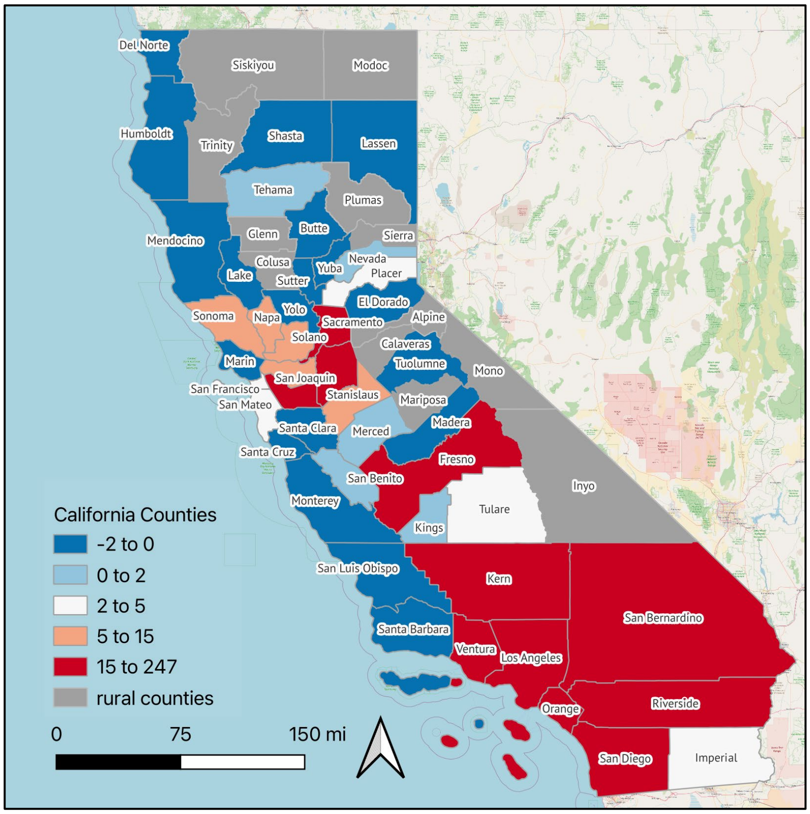
Figure 4 Location Quotient by MSA, 2014 and 2022



3.6 Subregional trends

We now turn to trends at the subregional level. First, we show gains and losses of W&D establishments by county from 2014 to 2022. See Figure 5. The grey areas represent rural counties with missing employment data. Blue shades are losses; white through red shades are gains. Note that losses are small; no county lost more than 2 establishments. In contrast gains are large; the maximum increase was 267. The largest gains are in Southern California, Fresno, and Sacramento, San Joaquin and Alameda counties. More modest increases are observed in the north of the greater San Francisco area. Counties in Northern California and along the central coast show little change.

Figure 5 Gains and Losses of W&D Establishments by County, 2014-2022



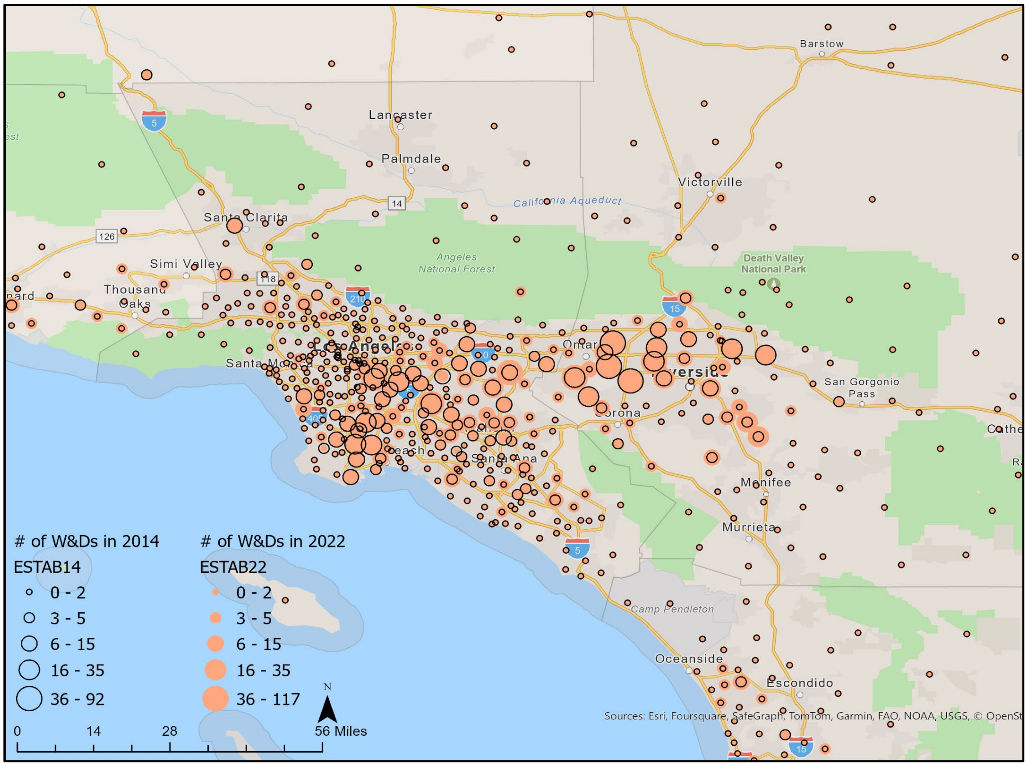
3.6.1 ZCTA subregional trends

We explore trends within metropolitan areas using the ZCTA data for W&D establishments. We show results for the Los Angeles, San Francisco and Sacramento, San Diego, and Central Valley regions. In each case we present two bubble maps. The first shows the number of W&Ds in 2014 and 2022. Black transparent bubbles represent 2014 and orange solid bubbles represent 2022. The legend scales are the same for each region. The second map shows changes in the number of W&Ds from 2014 to 2022. Dark blue is loss of 2 or more, light blue is stable (loss/gain +/- one), orange is gain of 2 to 5, and red is gain of over five. Few ZCTAs lost W&Ds, many remained stable, and some gained W&Ds. Thus, light blue is the dominant color on the map. All bubbles are located at ZCTA centroids.

Los Angeles Region

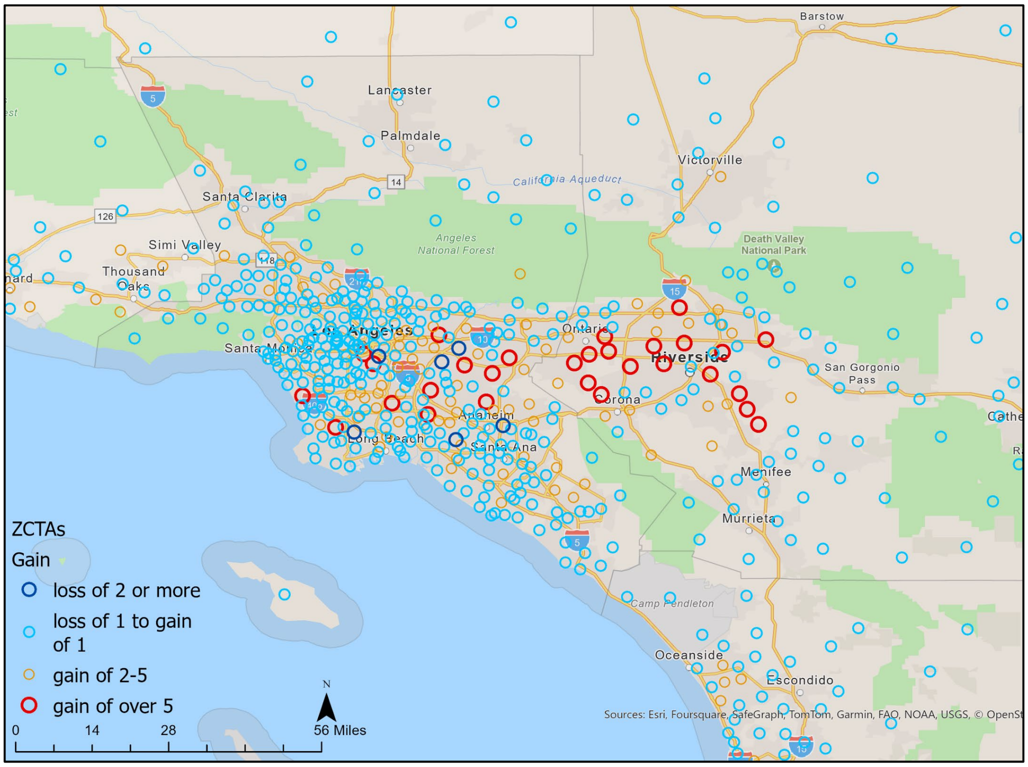
Figures 6 and 7 show results for the Los Angeles region. Figure 6 shows that W&Ds are concentrated in the heavily urbanized portion of the region (central part of Los Angeles County and Orange County) as well as along an east-west corridor from downtown Los Angeles to the Inland Empire. Development along the major highway corridors is evident (I-10, SR-60, I-15). A close look at the map reveals where increases are most evident – ZCTAs for which the orange bubble is notably larger than the black bubble. Note that this is evident throughout the region, including within the central core. Figure 7 shows gains and losses of W&D establishments. Dark blue bubbles (loss of 2 or more) are all in the urban core. In contrast there are many large red bubbles, representing gains of 5 to 28 new facilities. The gains are mostly concentrated along the I-10/SR 60/I-15 corridors, but there are also notable gains in the central core. The map suggests the reinforcement of existing spatial patterns.

Figure 6 Distribution of W&Ds in Greater Los Angeles Region, 2014 and 2022



Pacific Southwest Region UTC
University Transportation Center

Figure 7 Gains and Losses of W&Ds in Greater Los Angeles Region



San Francisco and Sacramento

Figures 8 and 9 give the same information for the greater San Francisco Bay Area and the Sacramento area. The spatial pattern is quite different from that of Los Angeles. W&Ds are tightly clustered around the San Francisco Bay, especially the east side which is the region's historic industrial zone. Others are scattered along the main freeways and around the Stockton area. W&Ds are distributed around the urban core of Sacramento and along the I-80 corridor. Close inspection of the bubbles reveals less evidence of W&D growth, and Figure 9 shows this more clearly. There are a few large red bubbles -- East Bay, Tracy, and Modesto, all near major highways – but most bubbles are light blue indicating little change. Dark blue bubbles appear in Sacramento and the SF East Bay (I-80) corridor. Overall, the maps show less growth than in the Los Angeles region, and clustering around the bay and Sacramento with a dispersed distribution in other portions of the region.

Figure 8 Distribution of W&Ds in greater San Francisco and Sacramento Regions, 2014 and 2022

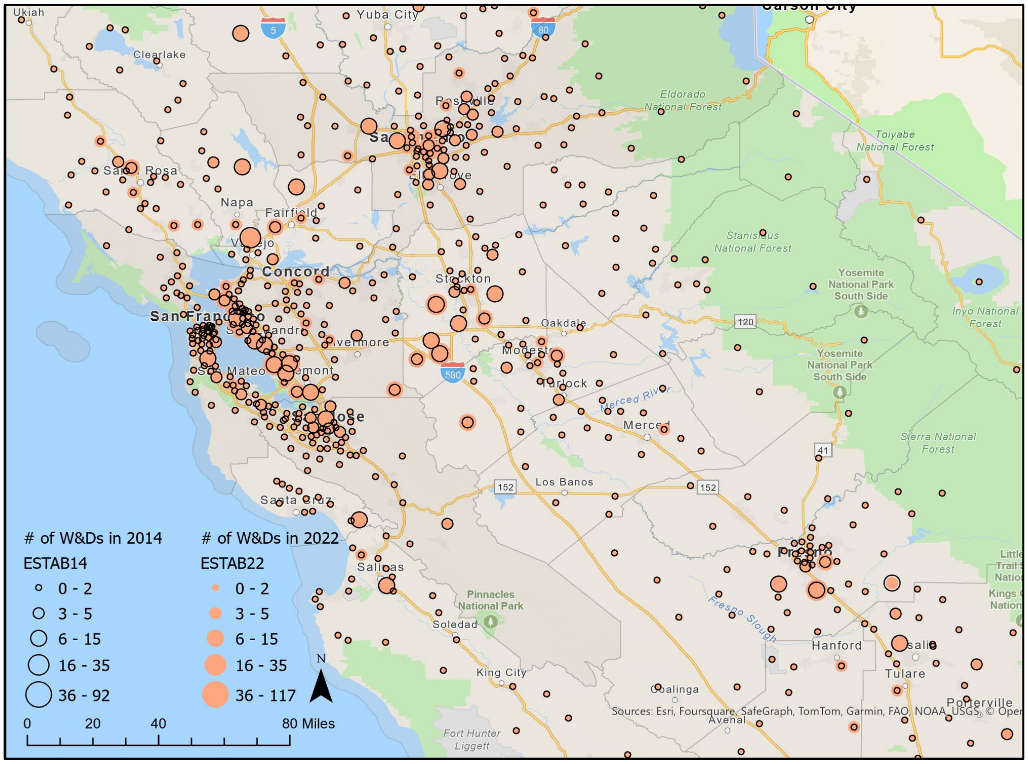
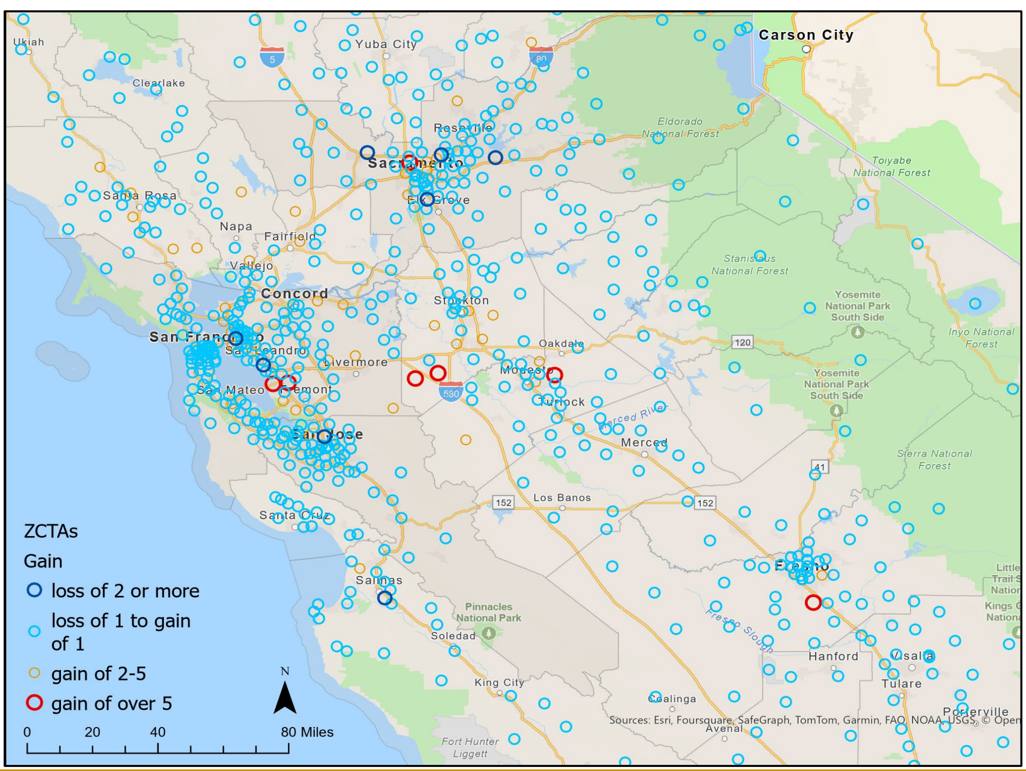


Figure 9 Gains and Losses of W&Ds in Greater San Francisco and Sacramento Regions



Pacific Southwest Region UTC
University Transportation Center

San Diego

Figures 10 and 11 show results for the San Diego area. The general pattern is more dispersed with a few concentrations at the national border, along I-5 in San Diego, and along SR 78 in the north. Much of the region to the east is national or state lands with little population. Figure 11 shows that gains took place at the border crossing and in the north county area, while the central area (most urbanized) remained stable with one loss of two or more in Chula Vista.

Figure 10 Distribution of W&Ds in Greater San Diego Area, 2014 and 2022

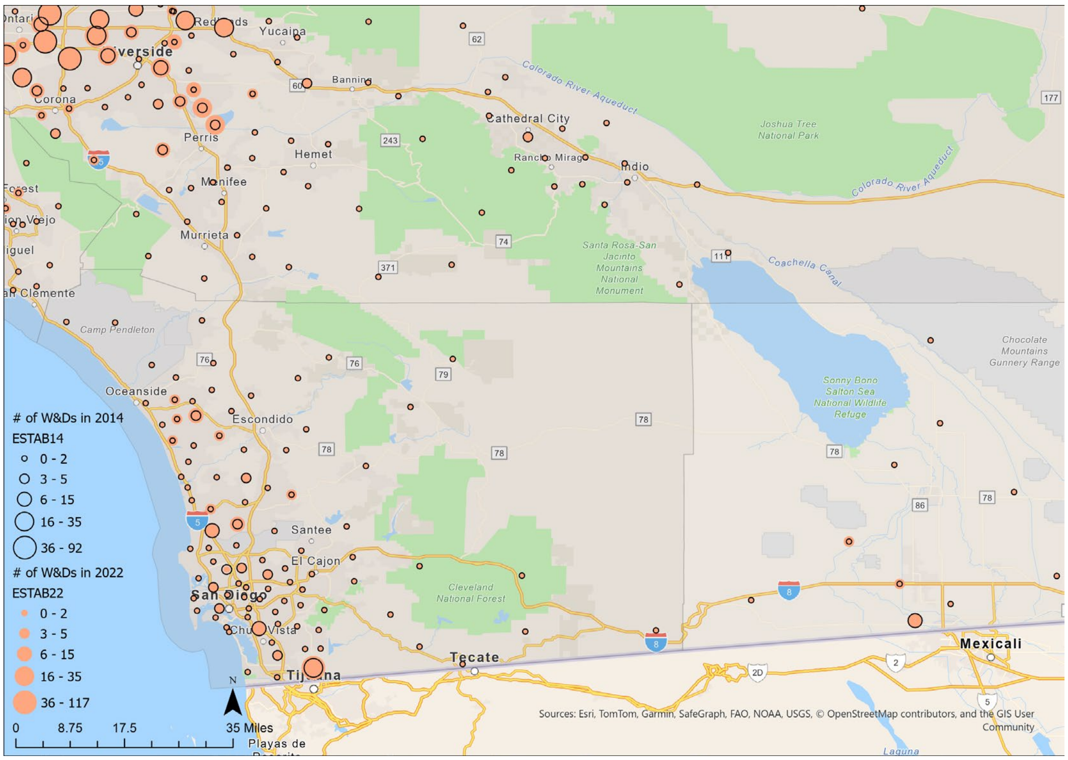
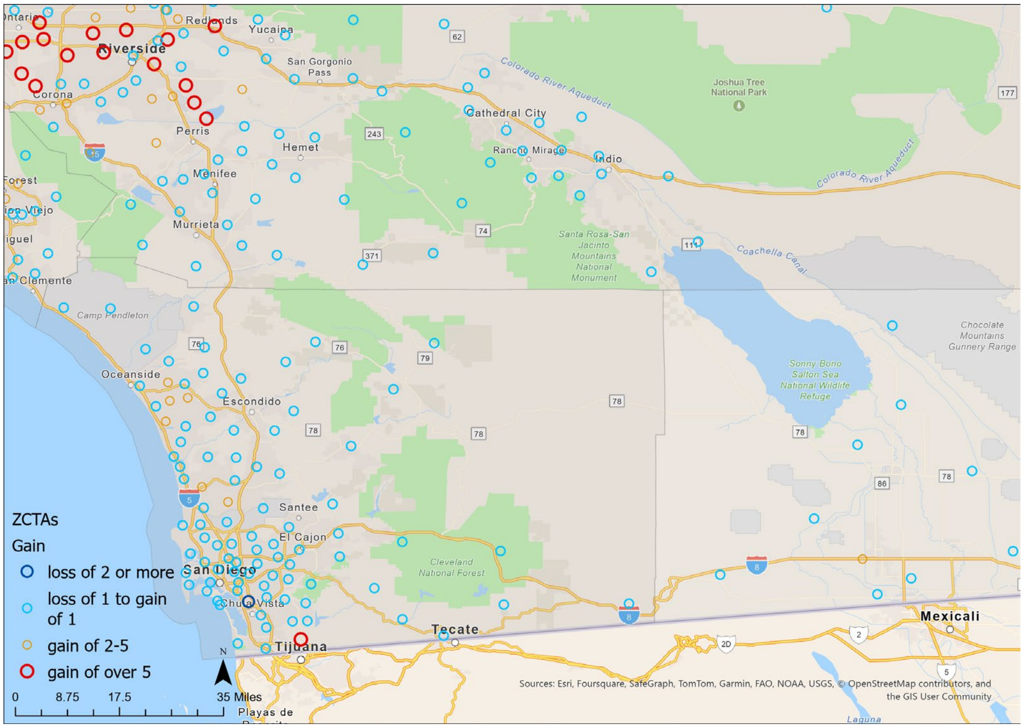


Figure 11 Gains and Losses of W&Ds in Greater San Deigo Area



Central Valley Region

Our final regional comparison is the Central Valley, from Fresno to Bakersfield. The largest concentrations are around Fresno and along the SR-99 corridor to Bakersfield. The overall pattern is quite dispersed. Figure 13 shows that gains have occurred in Fresno and near Bakersfield with the rest of the area remaining stable.

While W&D establishments grew 43% over our time period, our regional maps suggest that this growth has mostly reinforced existing spatial patterns. This contrasts with our previous study of trends 2003-2013 (Giuliano and Kang, 2017): in Los Angeles new clusters appeared in the Inland Empire and Santa Clarita; in San Francisco in Vallejo and the west bay; and in north San Diego. We further observe that while some losses took place in urbanized cores, so did many gains, which is consistent with demands related to increased e-commerce and short delivery times. Finally, while W&D clusters are clearly notable, these clusters exist within a larger context of dispersion. Some degree of W&D activity exists wherever there is sufficient population and economic activity.

Figure 12 Distribution of W&Ds in Central Valley Region, 2014 and 2022

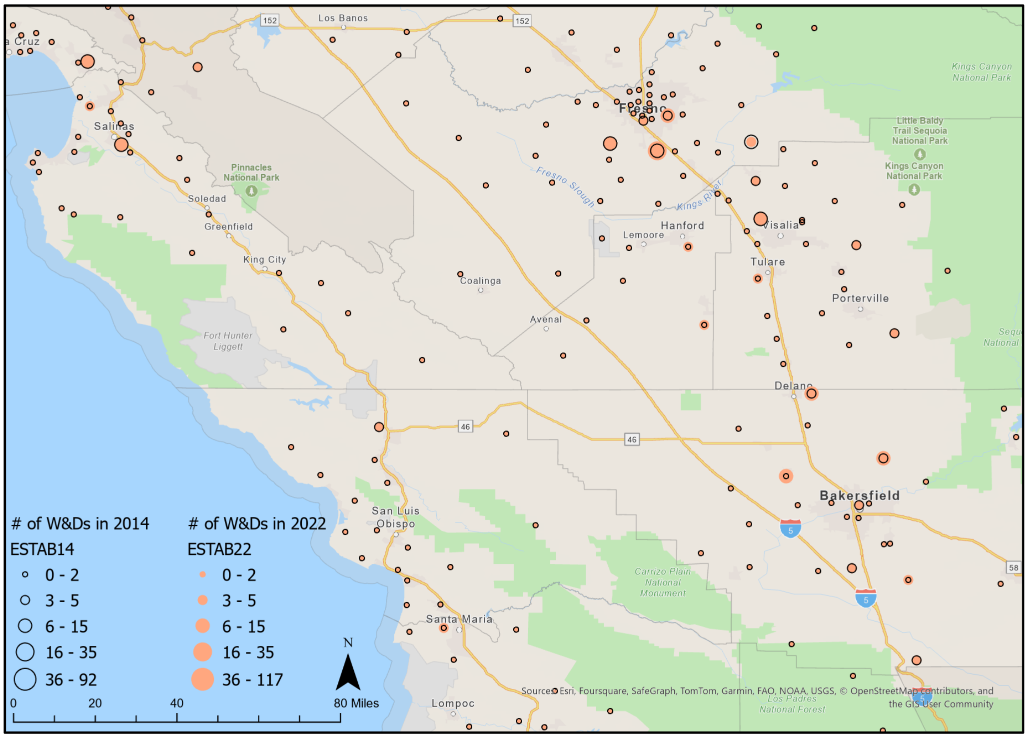
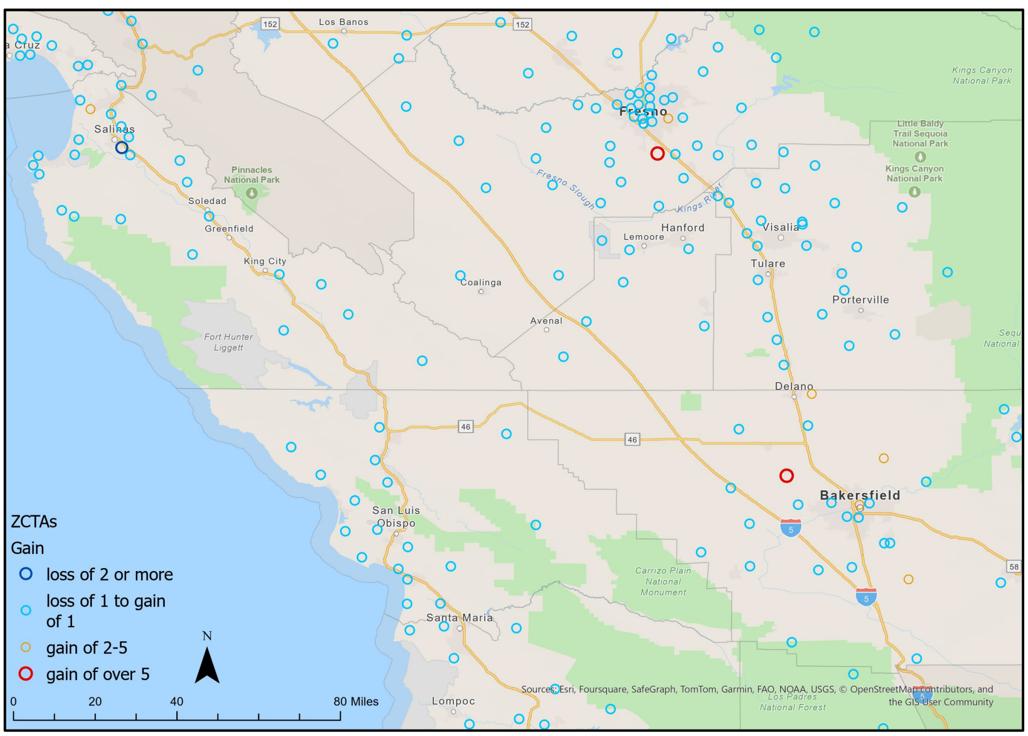


Figure 13 Gains and Losses of W&Ds in Central Valley Region



3.6.2 Centralization or decentralization

Lastly, we quantify the changes in W&D distribution patterns with a spatial measure. As discussed in our literature review, the decentralization of W&D activity has been of great interest. Literature from the past decade has demonstrated that decentralization is a global phenomenon among the largest metropolitan areas. The Los Angeles region in particular has been the subject of several studies. Are the trends we observe from 2014 - 2022 consistent with continued decentralization? We use the average distance from the central business district (CBD) to all W&Ds as our measure of change in spatial distribution. We calculate distance with respect to both establishments and employment. We define the CBD as the centroid of the ZCTA with the highest employment density of a metro area, and we use Euclidean distance. We test whether changes from 2014 to 2022 are statistically significant via Welch's t-tests on weighted averages. Average distance to the CBD is calculated as follows:

$$Y = \frac{\sum_j d_j e_j}{\sum e_j} \qquad (2)$$

Where,

Y = weighted average distance to CBD

d_j = distance from the CBD to ZCTA j (n; j = 1, 2, …, N)

e_j = number of W&D establishments or employment in ZCTA j

We show results only for the Level 1 areas for the following reasons: 1) the literature consistently show decentralization only for the largest metropolitan areas, and 2) smaller MSAs have rather low numbers of W&D, so any small change could generate a larger shift in average distance. Table 9 has two panels: the first gives results for establishments and the second gives results for employment. Within each panel the top row gives the percent change in establishments for each metro area. The changes range from an increase of 66% in Riverside to a 2% decline in San Jose. The next two rows give the average distance to the CBD in miles for 2014 and 2022 respectively. The last row gives the percent change in average distance. The percentage changes range from a decline of almost 10% (Riverside) to an increase of almost 13% (San Jose). We conducted weighted difference of means tests; none of the changes are statistically significant (results not shown). The second panel shows that the increase in employment was much greater than the increase in establishments; even in San Jose these was a 70% increase in W&D employment. The changes in average distance are smaller for employment, and none of the changes are statistically significant. It is rather remarkable that such enormous growth of the W&D industry has had little effect on the spatial distribution of W&D activity.

Table 9 Average Distance to CBD, 2014 and 2022, Six Largest MSAs

metro area	Los Angeles	Riverside	Sacramento	San Diego	San Francisco	San Jose
% change W&D establishments	49.4	65.8	16.0	48.4	37.0	-2.4
Ave distance to CBD 2014, establishments	16.32	16.85	11.96	14.28	18.71	7.97
Ave distance to CBD 2022, establishments	16.62	15.89	10.78	16.25	19.12	9.00
% Change Ave distance	1.84	-5.70	-9.87	13.8	2.19	12.92
% change W&D employment	66.3	159.3	71.6	92.0	205.6	70.0
Ave dist. to CBD 2014, emp	17.55	18.52	9.39	13.88	18.15	9.80
Ave dist. to CBD, 2022, emp	18.49	17.45	8.98	14.24	18.60	8.82
% Change Ave distance	5.36	-5.78	-4.37	2.59	2.48	-10.00

3.7 Conclusions on descriptive analysis

Our descriptive analysis leads to the following observations: 1) the W&D industry in California has grown much faster than the transport sector or the economy as a whole; 2) W&D activity is distributed approximately with total employment; the six largest metro areas in California account for about 83% of all jobs and 80% W&D jobs; 3) at the metropolitan level the relative shares of total employment have been stable over the period; the share of W&D employment in the largest MSAs increased slightly; 4) despite enormous growth, the spatial distribution of W&D showed little change: growth occurred both inside and outside urban cores, increasing the density of development.

Chapter 4: Understanding Trends

We now turn to explaining the trends observed in the previous chapter. We begin with our research framework, then describe our modeling approach and data, and finally present our results.

4.1 Research framework

W&Ds are part of a profit maximizing supply chain and will seek "productivity enhancing location attributes" (Sivitanidou, 1996, pp. 1262).[10] We assume that the observed W&D locations are a best proxy for optimal locations. Thus, we seek to explain why particular locations are attractive. Per the industrial location literature, important factors include land price, input costs (labor), transport costs, labor force access, market access and transportation access (Arauzo-Carod, et al. 2010). Our literature review (chapter 2) of warehouse location studies consistently identifies these factors as significant. Local land use policies also may play a role. Local governments may promote W&D development for economic growth or may see W&D development as an environmental problem and impose constraints. There is no readily available source for zoning and other local policies for the entire 2014-2022 time period of our data. We therefore do not directly incorporate land use policy in our model.

The general cross section model is:

$$W_i = f(L_i, M_i, A_i) \qquad (3)$$

Where,

W_i = Number of W&Ds in ZCTA i

L_i = vector of local market attributes of ZCTA i

M_i = vector of regional market attributes of ZCTA i

A_i = vector of transport access measures of ZCTA i

We define the local market as the ZCTA. Factors that would affect location at the ZCTA level include land availability and price, as well as labor force access. Population or employment density serve as proxies for land price, per the standard urban economics approach (Anas and Arnott, 1998). Density also serves as a proxy to land constraints. Labor force access is measured as the inverse-distance weighted population within 10 miles (the average commute distance) of the ZIP code centroid.

[10] W&Ds may or may not be built or owned by the firms that use them, but the principle holds in both cases. Firms that supply W&Ds would maximize profits by locating in places that are optimal for tenants.

The regional market is the MSA. Locations in metro areas that have more related industries or potential customers should be preferred. Regional market attributes include access to suppliers and linked industries (manufacturing, wholesale, and transportation), as well as to customers. We measure regional effects by access to linked industries: the share of linked employment relative to share of total employment. The third group of variables measures transportation access. These include distance to nearest airport, intermodal terminal, port and distance to nearest highways. We use the Euclidean distance from the centroid of a ZCTA to calculate distances. The major transport facilities tend not to be collocated, hence there is not a concern about multicollinearity.

Finally, it is possible that metropolitan size matters. There is an extensive literature on agglomeration economies – the external benefits to firms created by clustering.[11] These include sharing inputs and markets, matching labor and customers, and learning through knowledge sharing. In the case of W&D, clustering in large metro areas may allow for more efficient use of W&D facilities (e.g. by better balancing day to day demand). We use the metropolitan levels described in Chapter 3 as proxies for metropolitan size.

It is possible that the relative importance of these factors changes over time. As a metro area grows, density and land prices increase. Thus, W&D location may shift to lower density locations, trading off labor force or intermodal access for lower land price. Even without metropolitan growth, if scale economies increase demand for larger facilities, a similar shift to lower density locations could occur. If supply chains are increasingly national in scope, then attributes of the regional market may become less important. On the other hand, growing demand for short e-commerce delivery times may increase demand for access to the residential population. This suggests that the coefficients on our independent variables may be a function of the time period. If we observe changes in the coefficients, we have (indirect) evidence that time-related external factors are affecting location choice. It is possible to test for such changes by estimating cross sectional models for different years and formally testing for differences in coefficients between the time periods.

We have no priors regarding the temporal structure of independent variable effects. In our cross-section estimations, we are assuming that effects are contemporaneous. However, it is possible that effects are lagged. Once W&Ds are built, they remain in the stock for a long time, and markets may not be able to respond to shifts in demand immediately, given the length of the development process (shifts are more likely to be reflected in lease rates). We estimate time series models for a selected set of time intervals.

4.2 Modeling approach

We choose the number of W&D establishments in a ZCTA as our dependent variable. Although employment would be a better proxy for sector activity, the data problems described in

[11] See Duranton and Puga (2004), Rosenthal and Strange (2001, 2003), and Puga (2010) for the seminal work on agglomeration economies.

Chapter 3 preclude this option. There are some characteristics of our W&D data that must be considered in choosing appropriate model forms. First, our dependent variable is highly skewed and truncated at zero. Of the 1802 ZCTAs in California, about 70% had no W&D in any year. Among ZCTAs with at least one W&D establishment, the average number in 2014 was 5.01, the median was 3, and the maximum was 92. In 2022, the average increased to 5.08, the median remained at 3, and the maximum increased to 117. Transforming the variable to another form may address skewness, but not the truncation.

Second, our data are likely to be spatially correlated: if a given ZCTA has no W&D, it is likely that its neighbors also have no W&Ds. This is due to land use regulations and historical development patterns (e.g. residential areas are unlikely to have W&Ds). Spatial correlation can inflate the significance of explanatory variables that are themselves spatially correlated.

Third, our data are likely to be temporally correlated: if a given ZCTA has no W&D in 2014, it is unlikely to have one in 2015 or 2016. This is particularly true in cases where establishment data was imputed from historical data. This is not a problem for cross-sectional models but must be considered for time series models as will be further discussed below.

4.2.1 Cross section models

We use two model forms for cross section estimations. The first is a simple binary model to estimate the probability ($\Pr(y_i)$) of a given ZCTA having at least one W&D. The probability of a given ZCTA of having at least one W&D is:

$$\Pr(y_i) = \frac{exp^{v1}}{exp^{v1} + exp^{v2}}$$

$$and\ V_i = \alpha + X\beta + \varepsilon \qquad (4)$$

Where,

$y_i = outcome\ variable\ at\ ZCTA\ i$

$\alpha = intercept\ term$

$X = \text{vector of location factors with its systematic components } (\text{L}_i, \text{M}_i, \text{A}_i)$

$\beta = \text{vector of parameters to estimate by maximum likelihood}$

$\varepsilon = error\ term$

The binary logic model assumes error terms are independent and follow a standard logistic distribution. Binary logit is estimated via maximum likelihood estimation (MLE).

The second is a count data model to estimate the number of W&D establishments in a given ZCTA. The simplest count model assumes a Poisson distribution, meaning that the mean and variance of the dependent variable are equal. As noted above, this is clearly not the case for the distribution of W&D establishments. The variance is much larger than the mean, hence there is overdispersion in the distribution. The negative binomial model (NB) is a generalization of the Poisson model. It adds an additional term to account for the overdispersion. This term is approximately Gamma distributed. The probability of a given ZCTA having a given number of W&Ds is expressed as:

$$P(Y = y_i \mid \mu_i, \alpha) = \frac{\Gamma(y_i + a^{-1})}{\Gamma(a^{-1})\Gamma(y_i + 1)} \left(\frac{1}{1 + a\mu_i}\right)^{a^{-1}} \left(\frac{a\mu_i}{1 + a\mu_i}\right)^{y_i}$$

$$and\ \mu_i = \exp(X\beta) \quad (5)$$

Where,

y_i = outcome variable at ZCTA i

α = the variance parameter

$\Gamma(\cdot)$ = Gamma integral function

μ_i = the choice function

X = vector of location factors with its systematic components (L_i, M_i, A_i)

β = vector of parameters to estimate by maximum likelihood

As noted above, about 70% of all ZCTAs have no W&D establishments. The NB may not be a best fit for observations with a large share of zeros. An extension of the NB has been developed to account for this, the zero-inflated negative binomial (ZINB). The ZINB is a form of conditional probability. The first part is the zero-inflation model, which models the zeros in the data. It assumes that some zeros arise from a distinct process, meaning that there are certain observations that are guaranteed to be zero. The second part of the model estimates the counts for the observations that are not guaranteed to be zero. The ZINB is expressed as:

$$P(y_i) = \begin{cases} \pi + (1 - \pi) P_{NB}(Y = 0 \mid \mu, r), if\ y_i = 0 \\ (1 - \pi) P_{NB}(Y = y_i \mid \mu, r), if\ y_i > 0 \end{cases} \qquad (6)$$

Where,

$\pi \in [0,1]$ = the probability of an excess zero from the zero inflation process

$P_{NB}(Y = y_i \mid \mu, r)$ = the probability mass function of a negative binomial distribution

μ = the mean of the negative binomial distribution

r = the dispersion parameter

We thus have 3 models to estimate: binary logistic, negative binomial, and zero inflated binomial. We estimate the cross-sectional models for 2014, 2019, and 2022.

4.2.2 Controlling for spatial correlation

As discussed above and illustrated in Figures 3.6, 3.8, 3.10, and 3.12 W&Ds tend to be spatially clustered. We formally test for spatial correlation by conducting the Moran's I test on the dependent variable. Moran's I calculates the correlations between the value of the given variable (in this case number of W&D establishments) at a given location and that of its neighbors. Moran's I takes on values from -1 to 1; negative values imply negative correlations (neighbors have contrasting values), positive values imply positive correlations (neighbors have similar values). Calculating Moran's I requires defining what constitutes a "neighbor" and weights to represent nearness to neighbors. We constructed our neighbors list based on contiguous boundaries. Specifically, for each ZCTA, neighboring ZCTAs that share one or more boundary points are included. Equal weights are assigned to all neighboring ZCTAs in the list. We conducted global Moran's I tests for each year of the data; the statistic is positive and significant in every year. Results are given in Table 4.1 below. The magnitude of Moran's I is relatively low, suggesting that the degree of spatial correlation is relatively low. Over time the value of the statistic gradually increases, suggesting somewhat more clustering by 2022.

Table 10 Moran's I Statistics by Year

	Moran's I stat	Expectation	Variance	p-value
2014	0.24643	-0.0006	0.0002	<0.001
2015	0.24591	-0.0006	0.0002	<0.001
2016	0.24249	-0.0006	0.0002	<0.001
2017	0.26138	-0.0006	0.0002	<0.001
2018	0.25565	-0.0006	0.0002	<0.001
2019	0.26140	-0.0006	0.0002	<0.001
2020	0.27160	-0.0006	0.0002	<0.001
2021	0.27829	-0.0006	0.0002	<0.001
2022	0.30264	-0.0006	0.0002	<0.001

Spatial correlation can appear in regression models in two key forms. First, spatial correlation in the dependent variable occurs when the value of dependent variable in one location depends on the values in neighboring locations. To address this, a spatial lag term is added as an explanatory variable in a spatial lag model, with the spatial lag coefficient capturing the spatial correlation. For example, the binary model becomes:

$$Py_i = \frac{exp^{V_1}}{exp^{V_1} + exp^{V_2}}$$

$$and\ V_i = a + X\beta + wY\beta + \varepsilon \qquad (7)$$

where w is a spatial weight matrix and other terms are as defined for equation 4.

Second, spatial correlation may exist in the error terms, typically arising from unobserved spatial processes or omitted variables with a spatial structure. In such cases, a spatial error model incorporates a spatially lagged error term to account for the residual spatial autocorrelation, with the spatial error coefficient capturing the spatial correlation. Spatial dependence among explanatory variables, such as labor force access or employment density, is also common due to geographic or socio-economic factors. However, this does not inherently cause spatial correlation in residuals unless key spatially structured variables are omitted or the independent variables are mis specified. We have no reason to expect spatial error correlation in our models and there is no straightforward way to include spatial error lags in models that must be estimated via maximum likelihood (MLE), and all of our models are estimated with MLE. We therefore do not include spatial error lags in our models. We use the likelihood ratio, Wald statistic BIC (Bayesian Information Criterion) and AIC (Akaike's information criterion) to compare our three sets of models.

4.2.3 Time series models

Our cross-section models explain the location of W&D in a given year as a function of market and access conditions in the same year. W&D growth happens over time; the location of each new establishment is a choice based on the prior location choices of other W&Ds, linked industries, workforce, etc. It could therefore be argued that W&D locations in a given year are a function of locations in prior years. We have no theory regarding what type of temporal lag may exist – it could be one year, 10 years, or something in between. To simplify the problem, we use our target years of 2014, 2019, and 2022 as assume a first order auto-regressive model:

$$W_{i,t} = f\left(L_{i,t-k}, M_{i,t-k}, A_{i,t-k}\right) \qquad (8)$$

All terms are as defined as in equation (*3*). This model tests whether W&D locations in year *t* are a function of location characteristics in year *t-k*. In this case we have three different time period comparisons. As with the cross-section models, we estimate binary, NB and ZINB and use the same set of explanatory variables. We account for spatial lags as in the cross-sectional models, with the lags estimated for 2014 and 2019.

4.3 Data

Table 11 lists variables and definitions. The dependent variable was discussed above. Local labor market attributes include employment density and labor force access. Labor force access is measured as population of working age in the labor force within 10 miles of the ZCTA centroid, weighted by inverse distance. Share of linked industries includes transportation, manufacturing, and wholesale. Transportation access measures are drawn from publicly available map files. For distance to the nearest airport, we used the state's top 10 airports by cargo volume as reported by the Federal Aviation Administration (FAA). Due to the missing data problem explained in section 3.1.1, the descriptive statistics and the regression analysis excludes level 4. Thus, tables and figures starting from this section show results using level 1-3 MSAs.

Table 11 Variables and Definitions

Variable	Spatial unit	Definition
Dependent variable		
Binary ($W_{i,t}$) Count ($W_{i,t}$)	ZCTA ZCTA	W = 1 if ZCTA has at least 1 W&D, else = 0 W = number of W&D in ZCTA
Local market attributes		
Employment density	ZCTA	Jobs/mile2
Labor force access	ZCTA	Population ages 16 or older in the labor force within 10 miles of ZCTA centroid, weighted by inverse of distance
Regional market attributes		
Share of linked industries	MSA	Share of linked industry employment relative to total regional employment
Transportation access		
Distance to airport	ZCTA	Distance to nearest airport from centroid, miles
Distance to seaport	ZCTA	Distance to nearest seaport from centroid, miles
Distance to intermodal	ZCTA	Distance to nearest intermodal facility, miles
Distance to highway	ZCTA	Distance to nearest highway interchange, miles

Our data includes annual observations on W&D establishments and employment at the ZCTA level of geography from 2014 through 2022. The COVID pandemic began in 2020 and had a substantial impact on goods movement and supply chains as businesses closed or went remote and consumers shifted demand away from services and towards goods consumption. While public health restrictions were largely lifted by early 2021, impacts on the economy continued. We therefore split our series at 2019. The 2014 - 2019 period is one of general expansion of the economy. The COVID pandemic generated both a short economic recession and a large shock of increased goods demand which in turn generated increased demand for W&D space. The location factors in this latter period could well be different from the former period. To account for these possible differences, we use 2014, 2019, and 2022 as years for analysis and the intervals between these years as periods of analysis.

4.3.1 Dependent variable

As noted above, the dependent variable is highly skewed, as the majority of ZCTAs do not have W&Ds. Figure 14 gives the cumulative frequency distribution for the number of W&Ds for 2014, 2019, and 2022. The share of ZCTAs with at least one W&D is constant across the years, ranging between 32% to 34%. Figure 12 shows that the cumulative distribution is almost identical across the three years, another indication of the stability of the spatial organization of W&Ds.

Figure 14 Cumulative Distribution of W&Ds by ZCTA for level 1-3 MSAs, 2014, 2019, and 2022

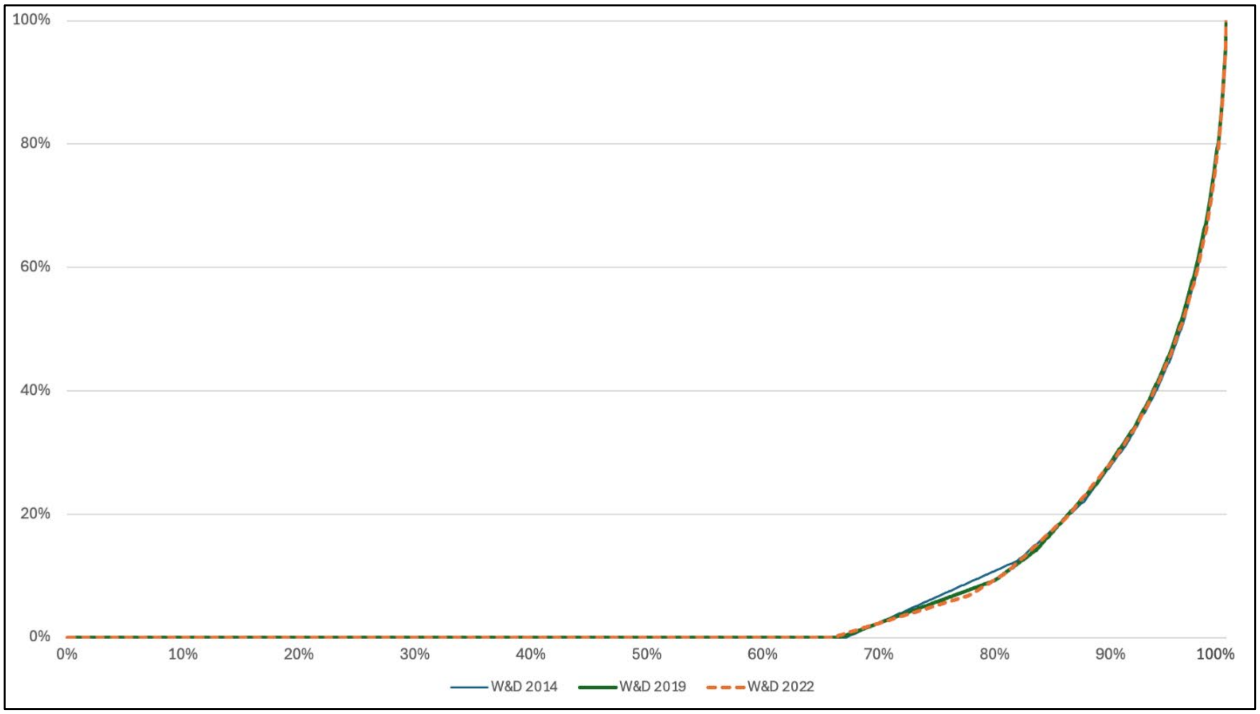


Table 12 compares presence or absence of W&Ds in each ZTCA across the comparison year. The horizontal panels compare 2014-2019, 2014-2022, and 2019-2022 respectively. More than 65% of all ZCTAs have no W&D in each comparison, the number of ZCTAs with at least one W&D increases slightly, and there are more ZCTAs that go from "no" to "yes" than from "yes" to "no". These observations show that relatively few ZCTA move from not having a W&D to having at least one W&D or move in the other direction. Growth in W&Ds has taken place where W&D activity previously existed.

Table 12 ZCTAs with at least one W&D Establishment, 2014-2019, 2014-2022, and 2019-2022

	Yes W&Ds in 2019	**No W&Ds in 2019**	**Total**
Yes W&Ds in 2014	476 (31.11%)	26 (1.70%)	502 (32.81%)
No W&Ds in 2014	34 (2.22%)	994 (64.97%)	1,028 (67.19%)
Total	510 (33.33%)	1,020 (66.67%)	1,530

	Yes W&Ds in 2022	**No W&Ds in 2022**	**Total**
Yes W&Ds in 2014	477 (27.2%)	25 (1.4%)	502 (32.81%)
No W&Ds in 2014	43 (2.4%)	985 (69.0%)	1,028 (67.19%)
Total	520 (29.6%)	1,010 (70.4%)	1,530

	Yes W&Ds in 2022	**No W&Ds in 2022**	**Total**
Yes W&Ds in 2019	510 (29.02%)	0 (0%)	510 (33.33%)
No W&Ds in 2019	10 (0.55%)	1,010 (70.42%)	1,020 (66.67%)
Total	520 (29.58%)	1,010 (70.42%)	1,530

4.3.2 Independent variable

Local and regional market variables

Table 13 gives descriptive statistics for the local market independent variables. Labor force population data was collected from American Community Survey (ACS) 5-year estimate data (S2301: Employment Status). We use population 16 years or older and in the labor force. We use LEHD (Longitudinal Employer-Household Dynamics) Origin Destination Statistics (LODES) for employment data, because of missing data at the two-digit sector level for small counties in the CBP data. The mean and median of both variables increase slightly over the period. Both variables are highly skewed; employment density is particularly skewed, reflecting the clustered nature of most employment.

Table 13 Descriptive Statistics, Local Market Variables

Variables	Mean	Median	SD	Min	Max
Labor Force Access 2014	69369	41384	76057	0	338729
Labor Force Access 2019	72471	43167	79344	0	352863
Labor Force Access 2022	72726	43711	78693	0	346426
Employment Density 2014	3521	418	24586	0	827400
Employment Density 2019	3785	477	22808	0	666100
Employment Density 2022	3923	484	23643	0	704011

Table 14 gives the regional shares of linked industries. We show values for each level and for each MSA in Level 1. A full table is available in Appendix 1. The table shows that average shares are quite stable over time and the average declines with each level, meaning that these industry sectors are a much smaller part of the local economy in smaller urban and rural areas. This stability contrasts with our previous study of 2003-2013, when linked industry share decreased across all MSA levels. Within the largest MSAs there are notable differences in industry mix. San Jose and Riverside have relatively high share of linked industries, likely due to high-tech manufacturing in San Jose. The Riverside share has increased over time, likely reflecting its continued growth in manufacturing and wholesaling. San Francisco, San Diego and Sacramento have notably lower shares, reflecting different industrial bases that are more service oriented.

Table 14 Share of Linked industry for each MSA Level and for Level 1 MSAs

Metropolitan level	2014	2019	2022
1	15.6%	15.4%	16.0%
2	14.5%	14.5%	15.2%
3	11.3%	11.5%	12.2%
Level 1 MSAs	**2014**	**2019**	**2022**
Los Angeles	17.8%	16.3%	16.0%
Riverside	18.3%	20.4%	23.4%
Sacramento	9.4%	9.3%	9.9%
San Diego	12.9%	13.2%	13.3%
San Francisco	12.8%	13.2%	13.2%
San Jose	21.0%	19.3%	19.6%

Transportation access variables

The transportation access variables do not change over time, as no new major facilities were built during our study period. Transportation facility locations were obtained from the California State GeoPortal. All distances are calculated as unweighted straight-line distance from ZCTA centroids to the given facility. The top 10 public airports by cargo volume as reported from FAA data were selected for airport access[12]. There are 11 seaports and 12 intermodal facilities included in the distance measures.[13] Lists of airports, seaports and intermodal facilities are available in Appendix 2-4. We use the California Enhanced National Highway System (NHS) network to measure distance from the nearest highway.

Table 15 gives descriptive statistics for the transportation access measures. The upper panel includes all ZCTAs' the lower panel includes only ZCTAs with at least one W&D. Figure 14 maps airports, seaports, intermodal facilities, and highways. As expected, most ZCTAs are far from airports, seaports and intermodal facilities, but relatively close to a highway. The distributions are all skewed, with the median much smaller than the mean. Access to all transport facilities is much greater in ZCTAs that have at least one W&D.

[12] Passenger Boarding (Enplanement) and All-Cargo Data for U.S. Airports, https://www.faa.gov/airports/planning_capacity/passenger_allcargo_stats/passenger/previous_years#2020

[13] Sources of location data: airports, https://gis.data.ca.gov/maps/082f0a402b354e53a7df995de3317fe2_0; seaports,https://gis.data.ca.gov/datasets/89282ed5837f4c779fabb082506b4528_0/explore?location=36.485334%2C-120.665196%2C6.94&showTable=true; intermodal, https://gis.data.ca.gov/datasets/5543e8b6b665463fb07b87cd8ba45043_0/explore?location=35.785313%2C-119.813850%2C7.76&showTable=true; highway, https://gis.data.ca.gov/datasets/1f71fa512e824ff09d4b9c3f48b6d602_0/explore?location=36.988180%2C-119.352150%2C6.77.

Table 15 Transportation Access Measures Descriptive Statistics, Distances in Miles

Variable	Mean	Median	SD	Min	Max
All ZCTAs					
Distance to airport	41.59	23.43	42.39	0.49	177.17
Distance to seaport	46.17	30.69	40.81	0.25	210.69
Distance to intermodal	44.44	32.24	40.13	0.52	228.13
Distance to highway	2.56	0.66	4.71	0.00	52.22
ZCTAs with at least one W&D					
Distance to airport	31.54	18.26	36.98	0.82	167.92
Distance to seaport	35.70	23.74	34.00	0.25	163.90
Distance to intermodal	34.99	22.21	35.39	0.52	197.69
Distance to highway	1.09	0.38	1.89	0.00	14.63

Figure 15 Transportation Access Facilities

Transportation access varies by metropolitan size, as shown in Table 16. For every type of facility, the largest MSAs have the highest level of access, and access declines systematically with level. The Level 1 ZCTAs have much greater accessibility – four times as much as level 2 for airports, and more than double for seaports and highways. Although we have no data to distinguish local or regional vs national or international activity, the lack of access to major facilities suggests that W&D activity in smaller MSAs and rural areas is locally oriented.

Table 16 Transportation Access Measures Descriptive Statistics by Metropolitan Level

Level	**Distance to airport**					**Distance to seaport**				
	Mean	Median	SD	Min	Max	Mean	Median	SD	Min	Max
1	21.24	16.49	19.29	0.49	174.29	30.96	22.97	28.27	0.48	210.69
2	81.57	78.78	46.85	3.30	177.17	74.68	69.12	46.19	0.25	182.67
3	88.28	86.53	43.65	10.41	170.73	86.05	84.46	38.14	18.1	168.40
Level	**Distance to intermodal**					**Distance to highway**				
	Mean	Median	SD	Min	Max	Mean	Median	SD	Min	Max
1	32.92	21.86	30.42	0.52	179.10	1.60	0.40	3.53	0.00	37.75
2	55.66	52.03	33.81	1.16	148.27	4.20	1.73	6.14	0.00	52.22
3	110.51	105.52	58.93	21.14	228.13	5.63	4.04	5.49	0.05	28.98

Finally, we look at highway access and W&D location. Table 17 shows the share of W&Ds located within one mile of the nearest highway by MSA size level. We show numbers for 2022 only, as the shares vary little across years. In level 1 MSAs, almost 90% of all W&Ds are within one mile of a highway. Large MSAs have the densest highway networks, hence a larger proportion of all the land area is within one mile of a highway than is the case for smaller MSAs. The share declines by about half at every level; for Level 3 MSAs, more than 75% of all W&Ds are *not* located within one mile of a highway.

Table 17 Share of W&Ds within One Mile of Nearest Highway, 2022

Level	**N**	**Share**
1	1891	89%
2	182	42%
3	20	24%
Total	2093	79%

4.4 Results

This section presents model results. We eliminate the Level 4 ZCTAs because 1) they account for just 1% of all W&D establishments. 2) a significant portion of the W&D data for Level 4 counties had to be imputed because of missing data problems after 2016. Thus there are more likely to be errors or outlier problems with the Level 4 ZCTAs.

4.4.1 Cross section models

We begin with the binary cross section models.

Binary models

Table 18 gives results for the binary model for 2014, 2019, and 2022. All the independent variables are in natural log form as discussed earlier. For each year we show results with and without a spatial lag for the dependent variable. We observe the following. First, there is clear evidence of spatial correlation in the residuals. Moran's I test for residuals is positive and significant in models without a spatial lag in all three years; when adding the spatial lag Moran's I becomes not significant and goodness of fit measures (AIC, BIC) improve with the spatial lag. Adding the spatial lag also affects the value of the coefficients of other independent variables. We therefore conclude that the spatial lag model is preferred.

Second, the local market variables, employment density and labor force access are significant and positive as expected. There is some indication of the effect of employment density increasing over time, while effect of labor force access remains stable. The coefficient for our regional market variable, share of linked industries, is not significant in any year.

Third, our transport access variable coefficients are not significant in any year. This is counter to expectations and implies that transport access is not a factor in W&D location, yet our descriptive statistics show that ZCTAs with at least one W&D have much greater access to highways, airport, seaports and intermodal facilities than ZCTAs without at least one W&D. This point will be further discussed after other model results are presented.

Fourth, the Level 1 MSA dummy variable coefficient is significant and negative with a value of about one, meaning that the likelihood of a given ZCTA having at least one W&D is much lower than the base case (Level 3). The Level 2 dummy coefficient is not significant in the spatial lag models. This result is explained by the geography of ZCTAs. There are many more ZCTAs in the large MSAs, and while a larger share of ZCTAs in the large MSAs have at least one W&D than is the case in the smaller MSAs, they also have the largest number of ZCTAs without at least one W&D, as shown in Table 19.

Table 18 Cross Section Binary Model Results, 2014, 2019, 2022, without and with Spatial Lag

	2014		2019		2022	
	Without spatial lag	With spatial lag	Without spatial lag	With spatial lag	Without spatial lag	With spatial lag
	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.
(Intercept)	-6.920 ***	-7.611 ***	-7.080 ***	-7.675 ***	-7.751 ***	-8.294 ***
Spatial Lag	-	1.553 ***	-	1.552 ***	-	1.506 ***
Local and Regional Market Attributes						
Employment Density	0.103**	0.114 **	0.118 **	0.136 ***	0.122 **	0.131 ***
Labor Force Access	0.623***	0.560 ***	0.599 ***	0.527 ***	0.627 ***	0.560 ***
Share of Linked Industries	-3.208	-2.829	0.231	0.429	2.518	2.527
Transportation Access Measure						
Distance to Airport	-0.096	-0.029	-0.074	-0.007	-0.016	0.044
Distance to Highway	0.118	0.209	0.063	0.153	0.054	0.129
Distance to Seaport	0.153	0.134	0.082	0.062	0.027	0.013
Distance to Intermodal Facilities	0.053	0.094	0.073	0.117	0.102	0.147 *
MSA size (base=level 3)						
Level 1 MSA	-1.215 ***	-0.932 ***	-1.279 ***	-1.038 ***	-1.336 ***	-1.097 ***
Level 2 MSA	-0.424	-0.318	-0.442 ***	-0.356	-0.515 *	-0.418
Goodness of fit						
AIC	1723.925	1657.428	1708.587	1657.532	1705.222	1657.445
BIC	1777.255	1732.092	1761.917	1716.195	1758.552	1716.108
Log Likelihood	-851.962	-825.714	-844.294	-817.766	-842.611	-817.723
Moran's I for residuals	0.134 ***	-0.007	0.136 ***	-0.008	0.132 ***	-0.007
# of obs	1530					

Table 19 Number of ZCTAs with and without W&D establishments by level

Year		Level 1	Level 2	Level 3	Total
2014	No W&D	660 (64.3%)	281 (72.1%)	87 (77%)	1028
	Yes W&D	367 (35.7%)	109 (27.9%)	26 (23%)	502
	Total	1027	390	113	1530
2019	No W&D	650 (63.3%)	281 (72.1%)	89 (78.8%)	1020
	Yes W&D	377 (36.7%)	109 (27.9%)	24 (21.2%)	510
	Total	1027	390	113	1530
2022	No W&D	641 (62.4%)	280 (71.8%)	89 (78.8%)	1010
	Yes W&D	386 (37.6%)	110 (28.2%)	24 (21.2%)	520
	Total	1027	390	113	1530

Negative binomial models

Tables 20, 21, and 22 give results for negative binomial models for 2014, 2019 and 2022 respectively. To show how model form affects results, we present four models: Poisson without spatial lag, negative binomial without spatial lag, negative binomial with spatial lag, and zero-inflated negative binomial with spatial lag. Goodness of fit measures are AIC and BIC. For the Poisson model we provide the dispersion measure which indicates the extent of overdispersion in the dependent variable. For the negative binomial models, we use Moran's I to test for spatial correlation. Our results are as follows.

First, in the Poisson model, coefficients of the independent variables are significant in all cases except share of linked industries in 2014, but the dispersion measure is very large and positive and goodness of fit measures relatively low. Clearly the Poisson model is not a good fit for our data. Second, as with the binomial models, Moran's I shows significant positive spatial correlation in the NB model; the NB model with spatial lag performs better. The goodness of fit measures are similar for the NB and ZINB models with spatial lags. We therefore focus on results for these models.

Employment density and labor force access variable coefficients are significant and positive in all cases for the NB model. Magnitude of coefficients changes little over the years, with labor force access having greater magnitude than employment density. Three of the four transport access variable coefficients are significant in 2014. Distance to airport is negative, meaning a probability of more W&Ds in any given ZCTA as distance declines. The distance to seaport and distance to intermodal facilities variable coefficients are positive, meaning the likelihood of more W&Ds increases with distance from these facilities, Most W&D activity is not directly related to international trade, and there are only a few seaports and intermodal facilities in the entire state. It seems reasonable that most W&Ds are located away from these facilities. In 2019 all four transport access measures have significant coefficients, but the highway coefficient has the wrong sign. In 2022 only distance to highway and to intermodal facilities have significant coefficients. There is some suggestion here that the effects of transport access vary across the years. As observed in the binary models, the Level 1 dummy variable coefficient is negative and significant with a value of about one. The Level 2 coefficient is not significant in any year. Results are consistent with the binary model.

The ZINB model is more complicated. The first column of coefficients is for the count portion of the model, conditional upon the ZCTA not having zero W&Ds. The second column of coefficients is for the probability of a ZCTA having no W&Ds. The spatial lag coefficient is relatively small and positive for the count estimation and larger and negative in the zero estimation. This makes sense: the more W&Ds a ZCTA's neighbors have, the less likely it is to have zero W&Ds.

The effects of local and regional market variables are different in each year. In 2014, labor force access is significant and positive, share of linked industries is negative and positive in the count estimation; none are significant in the zero-inflation estimation. In 2019 employment density is positive in the count model and labor force access is negative in the zero-inflation model. In 2022 none of the market variable coefficients are significant. These results are difficult to interpret, given the consistency of results in the NB and binary models.

Transportation access measures show similarly inconsistent results. In general, we would expect opposite signs between the count and zero inflation estimation, but most of the signs are the same, whether significant or not. The MSA dummy variable coefficients are more consistent. Level 1 is negative and significant in the count model, positive in the zero-inflation model for 2014 and not significant in 2019 and 2022. The magnitude of the Level 1 coefficient in the count model declines over time, which is not the case for the NB model. MSA level coefficients are never significant in the zero inflation estimations/ Finally, the log theta coefficient is significant in all years, indicating that the zero-inflation estimation has a significant negative impact on the count probability estimation. This provides statistical support for the ZINB form.

Table 20 Poisson, Negative Binomial, and Zero-Inflated Negative Binomial Model Results, 2014

	Poisson	Negative Binomial – without lag	Negative Binomial – with lag	Zero-Inflated Negative Binomial – with lag	
	Coefficients	Coefficients	Coefficients	Count model Coefficients	Zero-inflation model Coefficients
(Intercept)	-2.205 ***	-5.188 ***	-6.442 ***	-2.217 *	15.822 ***
Spatial Lag	-	-	0.263 ***	0.160 ***	-1.642 ***
Local and Regional Market Attributes					
Employment Density	0.063 ***	0.108 **	0.159 ***	0.062	-0.515
Labor Force Access	0.427 ***	0.662 ***	0.550 ***	0.366 *	-0.488
Share of Linked Industries	-1.442	-2.939	-3.452	-5.045 *	-10.130
Transportation Access Measures					
Distance to Airport	-0.635 ***	-0.541 ***	-0.276 ***	-0.311 ***	-0.349
Distance to Highway	0.134 **	0.158	0.196	0.243	0.108
Distance to Seaport	0.388 ***	0.338 ***	0.220 **	0.178 **	-0.068
Distance to Intermodal Facilities	-0.230 ***	-0.149 **	0.120 *	-0.148*	-1.630
MSA size (base=level 3)					
Level 1 MSA	-1.471 ***	-1.600 ***	-1.159 ***	-0.963 ***	0.247 ***
Level 2 MSA	-0.478 ***	-0.375	-0.236	-0.209	0.716
Dispersion Parameter (only for ZINB count model)					
Log (theta)	-	-	-	-0.727 ***	-
Goodness of Fit					
AIC	6486	3677.061	3578.553	3545.425	
BIC	6539.3	3735.724	3642.549	3668.085	
Log Likelihood	-3233	-1827.53	-1777.276	-1749.713	
Dispersion	12.109 ***	-	-	-	
Moran's I for residuals	-	0.223 ***	-0.015	-0.057	
# of obs	1530				

Table 21 Poisson, Negative Binomial, and Zero-Inflated Negative Binomial Model Results, 2019

	Poisson	Negative Binomial – without lag	Negative Binomial – with lag	Zero-Inflated Negative Binomial – with lag	
	Coefficients	Coefficients	Coefficients	Count model Coefficients	Zero-inflation model Coefficients
(Intercept)	-3.499 ***	-6.354 ***	-7.121 ***	-1.845	14.161 ***
Spatial Lag	-	-	0.232 ***	0.142 ***	-1.271 ***
Local and Regional Market Attributes					
Employment Density	0.096 ***	0.198 ***	0.229 ***	0.341 ***	0.229
Labor Force Access	0.412 ***	0.641 ***	0.532 ***	0.055	-1.193 ***
Share of Linked Industries	5.316 ***	1.540	-0.311	0.363	0.921
Transportation Access Measures					
Distance to Airport	-0.513 ***	-0.486 ***	-0.228 **	-0.180 *	-0.077
Distance to Highway	0.157 ***	0.218	0.242 *	0.422 ***	0.430
Distance to Seaport	0.323 ***	0.303 ***	0.172 *	0.122	0.095
Distance to Intermodal Facilities	-0.168 ***	-0.081	0.150 **	-0.079	-1.168 ***
MSA size (base=level 3)					
Level 1 MSA	-1.372 ***	-1.569 ***	-1.229 ***	-0.914 ***	0.549
Level 2 MSA	-0.491 ***	-0.332	-0.233	-0.255	0.373
Dispersion Parameter (only for ZINB count model)					
Log (theta)	-	-	-	-0.750 ***	-
Goodness of Fit					
AIC	7476.012	3912.848	3817.165	3777.145	
BIC	7529.342	3971.511	3881.162	3897.805	
Log Likelihood	-3728.006	-1945.424	-1896.583	-1864.573	
Dispersion	13.097 ***	-	-	-	
Moran's I for residuals	-	0.246 ***	0.001	-0.053	
# of obs	1530				

Table 22 Poisson, Negative Binomial, and Zero-Inflated Negative Binomial Results, 2022

	Poisson	Negative Binomial – without lag	Negative Binomial – with lag	Zero-Inflated Negative Binomial – with lag	
	Coefficients	**Coefficients**	**Coefficients**	**Count model Coefficients**	**Zero-inflation model Coefficients**
(Intercept)	-5.859 ***	-8.008 ***	-8.297 ***	-4.052**	10.217***
Spatial Lag	-	-	0.199 ***	0.110***	-1.088***
Local and Regional Market Attributes					
Employment Density	0.091 ***	0.213 ***	0.220 ***	0.098	-0.416
Labor Force Access	0.523 ***	0.694 ***	0.586 ***	0.368	-0.398
Share of Linked Industries	9.894 ***	5.578 ***	2.211	2.595	1.415
Transportation Access Measures					
Distance to Airport	-0.357 ***	-0.340 ***	-0.072	-0.110	-0.306
Distance to Highway	0.196 ***	0.226	0.236 *	0.332*	0.218
Distance to Seaport	0.293 ***	0.242 **	0.072	0.041	-0.023
Distance to Intermodal Facilities	-0.116 ***	-0.020	0.200 ***	-0.035	-0.977***
MSA size (base=level 3)					
Level 1 MSA	-1.262 ***	-1.454 ***	-1.123 ***	-0.634*	0.930
Level 2 MSA	-0.613 ***	0.362 ***	-0.211	-0.062	0.882
Dispersion Parameter (only for ZINB count model)					
Log (theta)	-	-	-	-0.752***	-
-					
AIC	8591.454	4139.937	4034.095	3981.965	
BIC	8644.784	4198.6	4098.092	4104.624	
Log Likelihood	-4285.727	-2058.968	-2005.048	-1967.982	
Dispersion	15.946 ***	-	-	-	
Moran's I for residuals	-	0.256 ***	-0.008	-0.052	
# of obs	1530				

Conclusions on cross-section models

We conclude the following from our cross-section estimations. There is significant spatial correlation in the dependent variable, and spatial lag models demonstrate better goodness of fit. Coefficients for local market variables are generally significant and have the expected positive sign. Our regional market measure is never significant in the binary and NB models.

Results for transport access measures are mixed. Coefficients for airport access are generally significant and of the expected sign. Positive signs for distance to seaports and intermodal facilities are explained by their geographic locations. Highway access is always of the wrong sign and significant in several cases. We empirically observe that nearly 80% of all W&Ds are within one mile of a highway and this proportion has actually increased slightly over time. One problem may be the correlation between the independent variables when in log form.

Regarding model forms, each of the spatial lag forms have similar values for the goodness of fit measures, despite differences in results across model forms. The ZINB results are somewhat less consistent across years and less comparable with the binomial and NB models for both market and access measures. The binomial and NB models suggest stability in spatial dynamics over the years.

We cannot directly compare the 2014-2022 results with the earlier 2003-2013 results because of differences in data, measures and model form. We can however make some general comparisons. Our results are generally consistent with local market measures; employment density and labor force access were consistently positive and significant in the previous research. Results for transport access measures are mixed. In the previous study distance to highway coefficients are generally significant and negative as expected (5 of 8 cases), contrary to the 2014-2022 results. For 2003-2013, distance to seaport is significant and positive in 7 of 8 model estimations, and distance to intermodal is significant and negative in 7 of 8 estimations. For 2014-2022, the coefficients for all but intermodal facilities are not significant in half or more of the estimations. Transport access does not seem to play as important a role in the later period.

4.4.3 Time series results

We estimate first order autoregressive models for three time period comparisons: 2014 -2019, 2019-2022, and 2014 - 2022 for the reasons discussed in the methodology section. We estimated binary, NB and ZINB without and with spatial lags. In all cases spatial dependence was demonstrated. We therefore discuss only the spatial lag models here.

Binomial time series models

Results for the binomial time series models are given in Table 23. Results are remarkably consistent with respect to goodness of fit, variable coefficient signs, magnitude and significance. They are also consistent with the cross-section results. This suggests a very stable temporal process: location factors for W&Ds have not changed, despite a more than doubling of the number of W&Ds over the period.

Table 23 Binomial Time Series Model Results, 2014-2019, 2019-2022, and 2014-2022

	2014 vs 2019	2019 vs 2022	2014 vs 2022
(Intercept)	-7.414 ***	-7.453 ***	-7.257 ***
Spatial Lag	1.570 ***	1.531 ***	1.557 ***
Local and Regional Market Attributes			
Employment Density	0.107 **	0.148 ***	0.119 **
Labor Force Access	0.577 ***	0.513 ***	0.563 ***
Share of Linked Industries	-3.357	0.352	-3.121
Transportation Access Measures			
Distance to Airport	-0.042	-0.003	-0.034
Distance to Highway	0.138	0.134	0.119
Distance to Seaport	0.107	0.051	0.093
Distance to Intermodal Facilities	0.066	0.102	0.057
MSA size (base=level 3)			
Level 1 MSA	-0.922 ***	-1.044 ***	-0.929 ***
Level 2 MSA	-0.251	-0.367	-0.266
Goodness of Fit			
AIC	1657.75	1658.969	1659.028
BIC	1716.413	1717.632	1717.691
Log Likelihood	-817.875	-818.484	-818.514
Moran's I for residuals	0.002	-0.005	0.003
# of obs	1530		

Negative binomial models

Results for the negative binomial models with spatial lags are given in Table 24 The models are consistent, but not as much so as the binomial models. Goodness of fit measures suggest a slightly better fit for 2014-2019, the interval before the COVID pandemic. The spatial lag is always significant and of similar magnitude. The 2019-2022 model results are a little different from the others: the coefficient for employment density is slightly higher, and for labor force access is slightly lower. Airport access is not significant and the other access coefficients are significant and positive. These subtle differences may reflect the effects of the COVID pandemic – the rapid increase in e-commerce and related demand for more urban warehouse space. Finally, the effect of MSA size is consistent across the comparison years. The coefficient for Level 1 is approximately one, as observed in the binomial time series models.

Table 24 Negative Binomial Time Series Results, 2014-2019, 2019-2022, and 2014-2022

	2014-2019	**2019-2022**	**2014-2022**
	Coefficients		
(Intercept)	-7.041 ***	-8.178 ***	-7.897 ***
Spatial Lag	0.273 ***	0.254 ***	0.292 ***
Local and Regional Market Attributes			
Employment Density	0.175 ***	0.203 ***	0.137 ***
Labor Force Access	0.616 ***	0.590 ***	0.674 ***
Share of Linked Industries	-4.122 *	1.871	-2.179
Transportation Access Measures			
Distance to Airport	-0.274 ***	-0.147	-0.221 **
Distance to Highway	0.197	0.223 *	0.161
Distance to Seaport	0.251 ***	0.162 *	0.259 ***
Distance to Intermodal Facilities	0.093	0.176 **	0.103
MSA size (base=level 3)			
Level 1 MSA	-1.164 ***	-1.080 ***	-1.004 ***
Level 2 MSA	-0.150	-0.190	-0.090
Goodness of Fit			
AIC	3823.657	4047.15	4063.335
BIC	3887.653	4111.146	4127.331
Log Likelihood	-1899.828	-2011.575	-2019.668
Moran's I for residuals	0.014	0.003	0.023
# of obs	1530	1530	1530

Zero-inflated negative binomial models

The last set of time series results are for the ZINB models. See Table 25. For each comparison year the spatial lag is significantly positive for the count estimation and significantly negative for the zero-inflation estimation, consistent with the cross-section results. Coefficient magnitudes are slightly different for 2019 - 2022. The local and regional market coefficients are also similar, with the exception of share of linked industries in 2014-2019. The transport access coefficients are consistent in most cases, and the MSA size coefficients are also consistent. As observed in the NB models, magnitude of coefficients is somewhat different for 2019-2022. Finally, the dispersion parameter is significant and of similar magnitude in each case.

Table 25 Zero-Inflated Negative Binomial Model Results, 2014-2019, 2019-2022, and 2014-2022

	2014 – 2019		**2019-2022**		**2014 - 2022**	
	Count model coef.	**Zero inflation model coef.**	**Count model coef.**	**Zero inflation model coef.**	**Count model coef.**	**Zero inflation model coef.**
(Intercept)	-2.887 *	12.495 ***	-3.817 **	10.661 ***	-3.426 **	11.330 ***
Spatial Lag	0.157 ***	-1.578 ***	0.141 ***	-1.265 ***	0.160 ***	-1.416 ***
Local and Regional Market Attributes						
Employment Density	0.048	-0.498 *	0.037	-0.561 **	-0.054	-0.649 ***
Labor Force Access	0.431 **	-0.413	0.402 **	-0.317	0.500 ***	-0.301
Share of Linked Industries	-4.849 **	-4.572	1.501	-1.158	-2.617	-1.323
Transportation Access Measures						
Distance to Airport	-0.310 ***	-0.402	-0.187 *	-0.261	-0.259 **	-0.326
Distance to Highway	0.220	0.029	0.253	0.063	0.160	-0.035
Distance to Seaport	0.216 **	0.053	0.127	-0.047	0.204 **	-0.079
Distance to Intermodal Facilities	-0.145 *	-1.182 ***	-0.065	-1.042 ***	-0.138	-1.035 ***
MSA size (base=level 3)						
Level 1 MSA	-0.849 **	0.542	-0.588 *	0.996	-0.563 *	0.853
Level 2 MSA	-0.036	0.928	-0.013	1.002	0.064	0.982
Dispersion Parameter						
Log (theta)	-0.787***		-0.811***		-0.829 ***	-
Goodness of Fit						
AIC	3785.602		3997.737		4012.337	
BIC	3908.262		4120.397		4134.996	
Log Likelihood	-1869.801		-1975.868		-1983.168	
Moran's I for residuals	-0.045		-0.041		-0.025	
# of obs	1530		1530		1530	

Pacific Southwest Region UTC
University Transportation Center

In terms of changes over time we conclude the following with regard to temporal relationships: lower employment density is associated with greater likelihood of zero W&Ds; higher labor force access is associated with greater likelihood of more W&Ds; distance to airport is (inversely) associated with more W&Ds; greater distance from seaports is associated with more W&Ds except for the 2019 - 22 interval; greater distance from intermodal facilities is associated with greater likelihood of zero W&Ds; and large MSA size is associated with fewer W&Ds, which is likely explained by the geography of ZCTAs as discussed in the cross section results.

Conclusions on time series models

Our conclusions on the time series models are similar to those for the cross-section models. Within each model form there is consistency in the results, with the possible exception of slight difference for the 2019-2022 comparison. Coefficients for local market measures are always positive and significant (for the ZINB significance for employment density is in the zero-inflation estimation). As with the cross-section models, transport access measures for highways and intermodal facilities are largely not significant; access to airport is negative and significant, and access to seaports is positive and significant in just over half of the model estimations. Our regional market access measure coefficient is significant in only two cases and has the wrong sign. Overall, transport access measures do not explain much of the variation in W&D location patterns.

As with the cross-section models, we provide a brief comparison between our 2014-2022 results and the previous 2003-2013 results. As with the cross-section models, local market factors are significant in all cases. Transport access measure coefficients are more consistently significant (only airport access is consistently not significant in the previous study) and have the expected signs. Spatial shifts were much more evident in the 2003-2013 period; W&Ds appeared in many new locations and decentralization was evident. It is possible that transport access factors played a greater role in the spatial expansion of W&D activity.

Chapter 5: Conclusions

Our research leads to the following general conclusions. First, the spatial organization of warehousing is remarkably stable: the industry sector more than doubled over our time period and the additional activity simply intensified the existing pattern. This spatial stability is illustrated in our region level maps of W&Ds, the demonstrated absence of change in location with respect to the CBD, and the general lack of significance of transport access variables in our statistical analysis. it appears that the decentralization or spillovers observed in the previous study have played out – those peripheral areas (e.g. Vallejo and I-80 corridor or San Joaquin Valley for San Francisco area, eastern Inland Empire or Bakersfield for Los Angeles area) are now part of the spatial pattern, but few new distant locations have emerged. This process of “infill” growth is consistent with e-commerce related demands for access to the population and short delivery times. It is also consistent with the increasing velocity of supply chains more generally. Our previous study gave other explanations for overall spatial stability: the concentration of population and jobs in a few very large metropolitan areas, the role of the largest metropolitan areas in the national and international economies, and path dependence driven by infrastructure investments and historical growth patterns. These explanations continue to hold.

Second, our results are not fully consistent with industrial location theory. Employment density, our proxy for land price and land scarcity is consistently significant, as is our measure of labor force access. Our measure for linked industries is generally not significant once spatial lags are introduced. We measure linked industries at the MSA level, and it is likely that the spatial lag captures this MSA level effect. Results for our transport access variables are not fully consistent. Access to airports is consistent, but access to highways is not; coefficients were generally of the wrong sign and not significant. Clearly highway access continues to be important: almost 80% of all W&Ds are located within 1 mile of the nearest highway. Coefficient signs and significance for access to seaport and access to intermodal facilities varied across model forms, hence no conclusions can be drawn regarding these measures.

Third, our time series models suggest that there may be differences in the spatial dynamics of warehouse location before and during/after the COVID pandemic. 2019-2022 was a period of very rapid W&D growth due to consumers shifts in buying patterns (both e-commerce and increased demand for goods) and associated supply chain bottlenecks and interruptions. Employment density had a slightly larger effect size which is consistent with increased demand for in-city warehouse space.

Our research has some limitations. First, the ZBP data are limited because we have only aggregated data: total numbers of W&Ds and employment at the ZCTA level. Because of censoring we cannot estimate establishments by size category, and we have no information regarding W&D function or physical attributes. Data for employment is missing in ZCTAs with few W&D facilities and consequently our statistical analysis is limited to establishments. We also have no information on location within the ZCTA. ZCTAs are geographically small in the cores of MSAs; in rural areas they are very large. Assuming that all activity is located at the centroid leads to ever larger potential errors as the geographic size of the ZCTA increases. An alternative to using ZCTA centroids is to assume some form of spatial distribution within the ZCTA, create a uniform overlay grid of smaller spatial units, and use the centroids of grid cells for distance measurement. In the absence of any data on how W&Ds might be spatially distributed within ZCTAs it is not clear that this approach would solve error problems.

Second, there are factors that influence W&D location not included in our analysis. Local zoning determines where new W&Ds can be built. Our models implicitly assume that land is available, and the market determines location choice. In past decades this was a reasonable assumption. Almost all municipalities have industrial zoning, and historically W&D location has been seen as positive for economic development. Environmental justice concerns have changed this perception. In recent years several municipalities have resisted W&D development, and W&D activity is facing increased regulation, as for example the South Coast Air Quality Management District's implementation of indirect source regulation on warehousing.[14] One indicator that unique local conditions are increasingly influencing W&D development would be if our models had less explanatory power over the time period. This is not the case. However, zoning and local regulations are an important topic for future research.

[14] https://www.aqmd.gov/home/rules-compliance/compliance/waire-program.

References

Allen, J., Browne, M., & Cherrett, T. (2012). Investigating relationships between road freight transport, facility location, logistics management and urban form. Journal of Transport Geography, 24, 45-57.

Cidell, J. (2010). Concentration and decentralization: the new geography of freight distribution in US metropolitan areas. Journal of Transport Geography, 18(3), 363-371.

Dablanc, L. (2019). E-commerce trends and implications for urban logistics. Urban logistics. Management, policy and innovation in a rapidly changing environment, 167-195.

Dablanc, L., & Ross, C. (2012). Atlanta: a mega logistics center in the Piedmont Atlantic Megaregion (PAM). Journal of transport geography, 24, 432-442.

Dablanc, L., Ogilvie, S., & Goodchild, A. (2014). Logistics Sprawl: Differential Warehousing Development Patterns in Los Angeles, California, and Seattle, Washington. Transportation Research Record: Journal of the Transportation Research Board, (2410), 105-112.

DeSousa, P. N., Ballare, S., & Niemeier, D. A. (2022). The environmental and traffic impacts of warehouses in southern California. Journal of Transport Geography, 104, 103440.

Dubie, M.D., Kuo, K.C., Giron-Valderrama, G., & Goodchild, A. (2020). An evaluation of logistics sprawl in Chicago and Phoenix. Journal of Transport Geography, 88, 102298.

Gingerich, K., & Maoh, H. (2019). The role of airport proximity on warehouse location and associated truck trips: Evidence from Toronto, Ontario. Journal of Transport Geography, 74, 97-109.

Giuliano, G., & Kang, S. (2018). Spatial dynamics of the logistics industry: Evidence from California. Journal of Transport Geography, 66, 248-258.

Giuliano. G, & Kang, S. (2017). Spatial dynamics of warehousing and distribution in California Final Report CA-17-2640. University of Southern California: METRANS Transportation Center.

Heitz, A., Dablanc, L., Olsson, J., Sanchez-Diaz, I., & Woxenius, J. (2018) Spatial patterns of logistics facilities in Gothenburg, Sweden. Journal of Transport Geography, 8, 102191.

Hesse, M. (2020). Logistics: Situating flows in a spatial context. Geography Compass.

Holl, A., & Mariotti, I. (2018). The Geography of Logistics Firm Location: The Role of Accessibility. Netw Spat Econ, 18, 337–361.

Isard, W. (1956). Location and the Space Economy. New York: John Wiley.

Jaller, M., & Pineda, L. (2017). Warehousing and Distribution Center Facilities in Southern California: The Use of the Commodity Flow Survey Data to Identify Logistics Sprawl and Freight Generation Patterns. UC Davis: National Center for Sustainable Transportation.

Jaller, M., Qian, X., & Zhang, X. (2020). E-commerce, Warehousing and Distribution Facilities in California: A Dynamic Landscape and the Impacts on Disadvantaged Communities. UC Office of the President: University of California Institute of Transportation Studies.

Jaller, M., Rivera, D., Harvey, J., Kim, C., & Lea, J. (2020). Spatio-Temporal Analysis of Freight Patterns in Southern California. UC Davis: Institute of Transportation Studies.

Jaller, M., Zhang, X., & Qian, X. (2022). Distribution facilities in California: A dynamic landscape and equity considerations. Journal of Transport and Land Use, 15(1), 755–778.

Jean-Paul, R. (2020). The distribution network of Amazon and the footprint of freight digitalization. Journal of Transport Geography, 88, 102825.

Kang, S. (2018). Warehouse location choice: a case study in Los Angeles, CA. Journal of Transport Geography, 88, 102297.

Kang, S. (2020). Relative logistics sprawl: Measuring changes in the relative distribution from warehouses to logistics businesses and the general population. Journal of Transport Geography, 83, 102636.

Kang, S. (2020). Why do warehouses decentralize more in certain metropolitan areas? Journal of Transport Geography, 88, 102330.

Losch, A. (1954). The Economics of Location. New Haven, CT : Yale University Press .

Moses, L. N. (1958). Location and the Theory of Production. The Quarterly Journal of Economics, 72(2), 259-272.

Onstein, A. T. C., Tavasszy, L. A., & van Damme, D. A. (2018). Factors determining distribution structure decisions in logistics: a literature review and research agenda. Transport Reviews, 39(2), 243–260.

Rivera-Royero, D., Jaller, M., & Kim, C.-M. (2021). Spatio-Temporal Analysis of Freight Flows in Southern California. Transportation Research Record, 2675(9), 740-755.

Rodrigue, J-P (2020) The distribution network of Amazon and the footprint of freight digitization, Journal of Transport Geography, 88, 102825, https://doi.org/10.1016/j.jtrangeo.2020.102825.

Sakai, T., Beziat, A., & Heitz, A. (2020). Location factors for logistics facilities: Location choice modeling considering activity categories. Journal of Transport Geography, 85, 102710.

Sakai, T., Kawamura, K., & Hyodo, T. (2019). Evaluation of the spatial pattern of logistics facilities using urban logistics land-use and traffic simulator Journal of Transport Geography, 74, 145–160.

Shearston, J.A., Johnson, A.M., Domingo-Relloso, A., Kioumourtzoglou, M.-A., Hernández, D., Ross, J., Chillrud, S.N., & Hilpert, M. (2020). Opening a Large Delivery Service Warehouse in the South Bronx: Impacts on Traffic, Air Pollution, and Noise. International Journal of Environmental Research and Public Health, 17(9), 3208.

Strale, M. (2020). Logistics sprawl in the Brussels metropolitan area: Toward a socio-geographic typology. Journal of Transport Geography, 88, 102372.

Todesco, P & Weidmann, U. (2016). Logistics Sprawl in the Region Zurich. In Proceedings of the 16th Swiss Transport Research Conference, Ascona, Switzerland.

Van den Heuvel, F., de Langen, P., van Donselaar, K., & Fransoo, J. (2013). Spatial concentration and location dynamics in logistics: the case of a Dutch province. Journal of Transport Geography, 28, 39-48.

Woudsma, C., Jakubicek, P., & Dablanc, L. (2016). Logistics Sprawl in North America: Methodological Issues and a Case Study in Toronto. Transportation Research Procedia, 12, 474–488.

Yang, Zhiwei, et al. (2022). Exploring location factors of logistics facilities from a spatiotemporal perspective: A case study from Shanghai. Journal of Transport Geography, 100, 103318.

Yuan, Q. & Zhu, J. (2019). Logistics sprawl in Chinese metropolises: Evidence from Wuhan. Journal of Transport Geography, 74, 242-252.

Yuan, Q. (2018a). Mega freight generators in my backyard: a longitudinal study of environmental justice in warehousing location. Land Use Policy, 76, 130–143.

Yuan, Q. (2018b). Environmental justice in warehousing location: State of the art. Journal of Planning Literature, 33.3, 287-298.

Yuan, Q. (2019a). Does context matter in environmental justice patterns? Evidence on warehousing location from four metro areas in California. Land Use Policy, 82, 328-338.

Yuan, Q. (2019b). Planning Matters: Institutional Perspectives on Warehousing Development and Mitigating Its Negative Impacts. Journal of the American Planning Association, 85(4), 525–543.

Yuan, Q. (2021). Location of warehouses and environmental justice. Journal of Planning Education and Research, 41.3, 282-293.

Data Management Plan

Products of Research

Our research is primarily based on two data sources produced by the US Department of Commerce, County Business Patterns (CBP) and Zip Code Business Patterns (ZBP). CBP provides annual county level data on employment and establishments at up to the 6-digit NAICS level. ZBP provides annual zip code level establishment and employment data up to the 2-digit NAICS level and establishment data at up to the 6-digit level. The data sources are publicly available for download. For spatial analysis the ZBP data are converted to Zip Code Tabulation Areas (ZCTAs). The ZCTA boundaries are provided by the US Census and are publicly available. We also used geographic data on California highways, airports, seaports, and intermodal facilities. The data are publicly available for download. Data on cargo traffic at California's airports were obtained from Federal Aviation Administration statistics, also publicly available.

Data Format and Content

The CBP and ZBP data are downloaded as CSV files, then processed and used to create Excel files for analysis. Infrastructure location data are downloaded and stored as QGIS files.

Data Access and Sharing

All data sources used in this report are publicly available. Websites for access are listed below:

CBP and ZBP data: https://www.census.gov/programs-surveys/cbp/data/tables.html
California Highway Data: https://gisdata-caltrans.opendata.arcgis.com/datasets/1f71fa512e824ff09d4b9c3f48b6d602_0/explore?location=36.855515%2C-119.352150%2C6.51
California Seaports:
https://gis.data.ca.gov/datasets/89282ed5837f4c779fabb082506b4528_0/explore?showTable=true
California Public Airport: https://gisdata-caltrans.opendata.arcgis.com/datasets/082f0a402b354e53a7df995de3317fe2_0/explore?location=36.945064%2C-119.334305%2C6.77
California Intermodal Facilities:
https://gis.data.ca.gov/datasets/5543e8b6b665463fb07b87cd8ba45043_0/explore?location=0.000000%2C0.000000%2C2.33
Passenger Boarding and All-Cargo Data for U.S. Airports:
https://www.faa.gov/airports/planning_capacity/passenger_allcargo_stats/passenger
LODES data: https://lehd.ces.census.gov/data/
Processed data files used in the research will be uploaded to Driyad

Reuse and Redistribution

Data cited or produced in this research have no restrictions on reuse and redistribution.

Pacific Southwest Region UTC
University Transportation Center

Appendix

Appendix 1 Share of linked Industries by MSA

Level	2014	2019	2022	MSA	2014	2019	2022
1	15.66%	15.41%	15.88%	Los Angeles-Long Beach-Anaheim	17.81%	16.27%	16.04%
				Riverside-San Bernardino-Ontario	18.29%	20.36%	23.37%
				Sacramento-Roseville-Folsom	9.44%	9.33%	9.86%
				San Diego-Chula Vista-Carlsbad	12.89%	13.24%	13.31%
				San Francisco-Oakland-Berkeley	12.82%	13.24%	13.15%
				San Jose-Sunnyvale-Santa Clara	20.98%	19.33%	19.63%
2	14.53%	14.50%	15.21%	Bakersfield	11.41%	10.91%	12.40%
				Fresno	13.81%	14.11%	14.66%
				Merced	19.88%	18.92%	17.50%
				Modesto	20.00%	18.88%	19.49%
				Oxnard-Thousand Oaks-Ventura	15.90%	13.84%	14.18%
				Salinas	8.15%	7.69%	7.51%
				San Luis Obispo-Paso Robles	10.97%	11.31%	11.22%
				Santa Cruz-Watsonville	12.20%	12.60%	13.96%
				Santa Maria-Santa Barbara	11.14%	10.41%	10.35%
				Santa Rosa-Petaluma	17.54%	17.15%	17.94%
				Stockton	20.19%	24.37%	28.89%
				Vallejo	14.62%	13.39%	14.47%
				Visalia	14.84%	15.51%	15.20%
3	11.25%	11.47%	12.16%	Chico	8.80%	8.78%	10.18%
				El Centro	8.94%	9.29%	9.41%
				Hanford-Corcoran	15.20%	17.91%	15.92%
				Madera	10.80%	10.15%	11.12%
				Napa	21.76%	21.94%	22.89%
				Redding	8.49%	9.07%	9.22%
				Yuba City	11.93%	11.60%	13.28%

Appendix 2 List of Airports

Airport Name	Airport ID
Los Angeles International	LAX
Ontario International	ONT
Metropolitan Oakland International	OAK
San Bernardino International	SBD
San Francisco International	SFO
San Diego International	SAN
Sacramento International	SMF
Sacramento Mather	MHR
Stockton Metropolitan	SCK
San Jose International	SJC

Appendix 3 List of Seaports

Seaport Name	County
Port of Redwood City	San Mateo
Port of Sacramento	Yolo
Port of Stockton	San Joaquin
Port of Richmond	Contra Costa
Port of Humboldt Bay Harbor	Humboldt
Port of San Diego	San Diego
Port of Oakland	Alameda
Port of San Francisco	San Francisco
Port of Long Beach	Los Angeles
Port of Los Angeles	Los Angeles
Port of Hueneme	Ventura

Appendix 4 List of Intermodal Facilities

Intermodal Facility Name	City	County
Oakland Intermodal Facility	Oakland	Alameda
ICTF Long Beach	Long Beach	Los Angeles
LA Transportation Center	Los Angeles	Los Angeles
LA Intermodal Facility	City of Commerce	Los Angeles
Santa Fe Hobart Yard	Los Angeles	Los Angeles
Lathrop Intermodal Facility	French Camp	San Joaquin
Stockton Intermodal Facility	Stockton	San Joaquin
Fresno Intermodal Facility	Fresno	Fresno
Railport Intermodal Yard	Oakland	Alameda
On-Dock Intermodal Facility	Wilmington	Los Angeles
City Of Industry Intermodal Facility	City of Industry	Los Angeles
San Bernardino Intermodal Facility	San Bernardino	San Bernardino

Pacific Southwest Region UTC
University Transportation Center