Working Paper

The Local Impact of Containerization

Leah Brooks
Trachtenberg School of Public Policy and Public Administration
George Washington University

Nicolas Gendron-Carrier
Department of Economics
University of Toronto

Gisela Rua
Division of Research and Statistics
Board of Governors of the Federal Reserve System

Friday 9th September, 2016
Preliminary. Please do not cite without authors' permission.

Thanks to Elliot Anenberg, Nathaniel Baum-Snow, Paul Carrillo, Jonathan Dingel, Marco Gonzalez-Navarro, Jessie Handbury, Thomas Holmes, Peter Morrow, Justin Pierce, Jordan Rappaport, Daniel Trefler, and Matthew Turner for many helpful comments and discussions. We are also very appreciative to research assistants James Calello, Adrian Hamins-Puertolas, Arthi Rabbane, Alex Severn, and Daniel Walker who helped with data entry. We are grateful to Matthew Turner and Gilles Duranton for sharing the 1956 County Business Patterns data. Finally, we thank Nora Brooks, retired reference librarian and Leah's mother, whose combination of digital and pre-digital skills uncovered the 1948 port-level trade data.

The Local Impact of Containerization

Abstract

This paper exploits the advent of containerization, a technological shock that dramatically reduced international shipping costs, to examine how access to international markets affects local economic growth. We contend with the non-random adoption of containerization by employing a novel instrument: being near a very deep port in 1953, before containerization. Because container ships are much larger and displace more water than their predecessors, they require substantially deeper ports. Despite their value in the post-container era, very deep ports had no particular competitive advantage before the advent of containerization. Analogous to a cost shifter, port depth should affect the supply of ports, but have no effect on the demand for ports. To estimate the impact of containerization on local economic activity, we construct a county level panel dataset describing the evolution of population, employment, port facilities, and other variables in the United States from 1950 to 2010. Consistent with the predictions of a standard new economic geography model, we find substantial increases in population, employment, and wages in U.S. counties near containerized ports. Population gains are roughly equal to the average increase in population for all counties from 1950 to 2010. Also consistent with the model, we find that containerization-induced population growth dissipates as distance to the containerized port increases. Finally, containerization-induced population growth is centered primarily in counties with initially weaker manufacturing sectors.

Leah Brooks
Trachtenberg School of Public Policy
George Washington University
805 21st St. NW, Room 601F
Washington, DC 20006
lfbrooks@gwu.edu

Nicolas Gendron-Carrier
Department of Economics
University of Toronto
150 St. George Street
Toronto, ON, Canada, M5S 3G7
nicolas.gendron.carrier@mail.utoronto.ca

Gisela Rua
Division of Research and Statistics
Board of Governors of the Federal Reserve System
20th and C Streets, Mail Stop 82
Washington, DC 20551
gisela.rua@frb.gov

Despite a vast and prominent theoretical literature that emphasizes the gains from market integration (e.g. Krugman, 1991; Fujita et al., 1999; Allen and Arkolakis, 2014; Redding, 2016), causal evidence on whether and how reductions in trade barriers increase economic activity is limited. In this paper, we use the advent of containerization, a technological shock that dramatically reduced international shipping costs, to examine how access to international markets affects local economic growth.

Containerization is premised on a simple insight: Packaging goods for waterborne trade into a standardized container makes them cheaper to move. Containerization simplifies and speeds packing, transit, pricing, and the transfer from ship to train to truck; it also limits previously frequent and lucrative pilferage. From the advent of international container trade in 1966 to 1981, Bernhofen et al. (2016) estimate that containerization caused international trade to grow by more than 1,000 percent. Containerized trade now dominates ocean shipping, and containers account for well over 75 percent of U.S. domestic rail traffic (Rodrigue, 2015).

To understand containerization's impact on local economic growth and guide our empirical investigation, we turn to a standard new economic geography model based on Helpman (1995), Redding and Sturm (2008), and Redding and Turner (2015). In such a model, the spatial distribution of economic activity is determined by the trade-off between centralizing and decentralizing forces. Centralizing forces include economies of scale in production, love of variety in consumption, and costly transportation. Decentralizing forces include land rents and commuting costs, both of which increase with population and are a source of congestion. A reduction in transportation costs—such as that yielded by containerization—causes increases in population, employment, and wages in areas closest to the transit cost advantage. Because the transit cost advantage declines with distance to a containerized port, the model also predicts that containerization's impact on key economic outcomes should attenuate with distance to a containerized port.

To empirically establish the causal effects of containerization on local economic growth, we must contend with the non-random adoption of containerization. We address the endogeneity of proximity to containerized ports with a novel instrument: proximity to a very deep port in 1953, before containerization.

The first requirement for a good instrument is that it is correlated with the endogenous variable. Container ships are substantially larger than their predecessors, and displace more water. They therefore require deep ports. While a port can be arbitrarily deep in the absence of cost concerns and environmental regulations, initially deeper ports are cheaper to convert to containerized ports because they require less dredging. This instrument is thus analogous to the cost shifter instruments used in the industrial organization literature. Empirically, we find a very strong relationship between the instrument and the endogenous variable.

The second requirement is that proximity to a very deep port in 1953 affects local economic growth only through its impact on containerization. Although ports varied in depth before containerization, being a very deep port—beyond 25 or 30 feet—posed no particular competitive advantage. Most ships did not displace enough water to require more depth. Crucially, being a very deep port matters only after the invention and diffusion of containerized shipping. Thus, we parametarize our instrument as a county's proximity to a very deep port, where the depth cut-off is beyond what was generally considered a useful depth in the pre-containerization era. To validate the instrument, we show that the identifying variation in the instrument is unrelated to plausible covariates of concern.

Our causal estimates of the impact of containerization on local economic growth rely on the quasi-random variation in initial depths. The estimates compare counties that are treated with a nearby containerized port—because they had nearby ports that were very deep before the advent containerization—to otherwise similar counties.

To undertake the estimation, we bring together data from a variety of sources on population, employment, wages, and port facilities for a panel of US counties from 1950 to 2010. From the Census, we use data on population and income. We rely on County Business Patterns (starting in 1956) for data on employment and payroll by industry. We combine these data with port level data on location, depth, and size, as well as measures of pre-containerization international trade flows. We observe the year of first containerization of any port in every county.

Using these data, our OLS and IV estimates show that proximity to a containerized port increases local economic growth. Our most complete instrumental variables estimates find that counties within 100 km of a containerized port experience a statistically significant 47 percentage points increase in population growth, or about 125 percent of the average increase from 1950 to 2010. We find that containerization's impact on population growth attenuates with distance to a containerized port. We also find increases in employment and nominal wages that follow a similar geographic pattern.

These results are robust to limiting the sample to only port-proximate counties, and to additional controls for pre-existing population levels. To assess whether these results are unique to the US, we also use a sample of world cities to do a similar analysis on population. In this world sample, we again find that proximity to a containerized port causes increases in population growth, though of smaller magnitude.

Finally, we consider whether containerization-induced population gains are spread evenly across treated counties, or whether they are concentrated in counties with particular initial conditions. We find that containerization's impact on population growth is concentrated in counties with initially weaker manufacturing sectors. Income gains accrue disproportionately in counties with higher than median initial education.

Our work makes two main contributions. First, we introduce a new class of instrument to the literature investigating the effect of transportation infrastructure on the growth of cities and regions (e.g. Baum-Snow, 2007; Duranton and Turner, 2012; Donaldson, forthcoming). Past papers in this literature have either relied on a structural model for identification or have used planned or historical routes as instruments. Papers that follow a structural approach include Donaldson (forthcoming), Donaldson and Hornbeck (2016), and Allen and Arkolakis (2014). Baum-Snow (2007) pioneered the planned route IV and Duranton and Turner (2012) pioneered the historical route IV. [1] Our instrument is most analogous to a cost shifter instrument, as is popular in the industrial organization literature.

Our second contribution is our analysis of the local impact of international trade. Pascali (2014) analyzes the impact of a decline in international transportation costs on country growth. Duranton and Turner (2012), Donaldson (forthcoming), and Donaldson and Hornbeck (2016), among others, analyze the impact of a reduction in inter-regional trade costs on local economic growth. Our paper contributes to this literature by directly considering how a large decline in international transportation costs affects local economic growth.[2]

The remainder of this paper is organized as follows. The following section provides background on containerization, Section 3 outlines the theoretical framework, and Section 4 discusses the data. We present empirical methods in Section 5, and results in Section 6. We conclude with Section 7.

[1]See Redding and Turner (2015) for a recent survey of the literature.

[2]Our paper is also related to a growing literature in international trade that looks at the impact of trade on local labour markets (e.g. Topalova, 2010; Autor et al., 2013; Dix-Carneiro and Kovak, 2015). These studies suggest that trade can have substantial localized effects.

2 Containerization

In this section we discuss the rise of containerization, its impact on transportation costs, and the factors that drive containerization's geographic pattern of adoption.

Before goods went into the box, shipping was expensive and slow. Vessels spent weeks at ports while gangs of dockworkers handled cargo piece by piece. Port costs accounted for a sizeable share of the total cost of the movement of goods. The American Association of Port Authorities estimated that in-port costs, primarily labor, accounted for half the 1960 cost of moving a truckload of medicine from Chicago to Nancy, France (Levinson, 2008, p. 9).

In response to these high costs, producers searched for alternatives. Trucker and entrepreneur Malcolm McLean is generally credited with being the first to match vision with reality when he moved 58 truck trailers on a ship from Newark to Houston in 1956 on the maiden container voyage.

Despite this small-scale start, container technology diffused extremely rapidly across the United States. The bulk of domestic containerization adoption occurred in the 1960s, as shown in Figure 1a, which reports the total number of US containerized ports by year. Adoption continued at a slower pace throughout the 1970s and 1980s and plateaued therafter.

Adoption of containerization in the rest of the world followed a similar pattern, roughly one decade delayed. Figure 1b shows that the majority of non-US containerization occurred in the 1970s (see also Rua (2014)). The pace of adoption in the US and across the world is consistent with the initial pattern of containerized trade. Until at least the mid 1960s, containerized trade was primarily domestic. The first international container service did not begin until 1966, nearly a decade after the first US shipment.

Containerized trade is now central to the global economy. Bernhofen et al. (2016)

estimate that containerization caused international trade to grow by more than 1,000 percent over the 15 years following 1966. As of 2013, containerized trade accounted for over half of global non-commodity trade (United Nations Conference on Trade and Development, 2013).[3]

Containerized trade relies on two key innovations. The first is the mechanization of container movements. Rather than workers with carts, specialized container cranes lift containers in and out of ships, around the port, and onto rail cars and trucks. This mechanization substantially decreased per unit labor costs, cut time in ports and made ever-larger ships viable. Today's Post-Panamax ship is more than 17 times larger than the first ship to carry container goods in 1956. In short, containerization innovations yielded tremendous economies of scale in waterborne trade.

The second key innovation of containerized trade is the development of common standards for container size, stacking techniques, and grip mechanisms. These standards allow a container to be used across modes of transportation—ships, trucks, rail—and across countries. The U.S. standard for containers was adopted in the early 1960s, and the international standard followed in the late 1960s, as indicated in Figure 1.

To achieve economies of scale, containerization requires physical changes to ports. In breakbulk ports, ships pulled into finger piers and workers on- and off-loaded items. Ports were centrally located within cities and used a large amount of labor and a moderate amount of land for warehousing and storage. In contrast, containerized ports require substantially less labor per unit of weight and a rather large amount of land. This land is both for the large cranes that move containers, and for the marshalling of containers and trucks (Rua, 2014).

These requirements for ports have shifted the distribution of dominant ports. Of the

[3]While containers are appropriate for carrying many goods, as diverse as toys and frozen meat, some goods are not yet containerizable. Both "non-dry cargo" and "dry-bulk commodities" such as oil, fertilizers, ore, and grain cannot be shipped inside "the box."

ten largest ports before containerization (in 1955, measured in terms of international trade), two never containerized: New York (Manhattan), NY and Newport News, VA. In fact, the Port of Manhattan, the largest in the world in 1956, no longer exists as a freight port. Of today's 25 largest ports, four did not rank in the pre-containerization top 25. Only two of the modern ten largest ports were in the pre-containerization top ten: Norfolk, VA and Los Angeles, CA.

The adoption of container technology at ports is driven by demand for port-shipped goods, by the political constraints on port construction, and on the natural characteristics of a port that determine port construction costs. On the demand side, we anticipate that containerization is driven by the initial market size around the port and the expectation of gains due to increased trade. For example, West Coast port containerization may have been led in part by anticipated growth in trade with Asia.

On the supply side, it is easiest to construct a containerized port where there is sufficient available land and an adequately deep harbor. We discuss the depth issue at length later, but suffice it to say that since container ships carry more volume, they ride deeper in the water than their predecessors. These ships therefore require deeper ports than breakbulk ships. We thus argue that pre-containerization harbor depth serves as a cost shifter in the supply of containerized ports.

On net, the literature finds that containerization substantially decreased the cost of waterborne trade. While Bridgman (2014) and Hummels (2007) note only a small decline in shipping rates, there are many reasons to believe that traditional measures of shipping costs understate the true cost advantage yielded by containerization. First, sources widely agree that containerization cuts the time ships spend at port and thus the total time in transit. Hummels and Schaur (2013) estimate that each day in transit is worth between 0.6 to 2.1 percent of the value of the good, showing that the time benefits of containerized shipping are non-negligible. Second, losses to pilferage plummeted with containerization. Wilson (1982) estimates loses to pilferage at roughly 25% in the break-bulk era, and near zero in the container era.[4] Finally, containers also protect goods from unintentional damage and allow different kinds of goods, with different destinations, to be shipped together.

3 Theoretical Framework

In this section, we outline a standard new economic geography model to illustrate containerization's impact on local economic growth based on Helpman (1995), Redding and Sturm (2008), and Redding and Turner (2015).[5] The model is characterized by the trade-off between the centralizing forces that tend to group people and firms in a limited number of locations and the decentralizing forces that tend to disperse them. Centralizing forces include economies of scale in production, love of variety in consumption, and costly transportation. Decentralizing forces include land rents and commuting costs, both of which increase with population and are a source of congestion. We use this framework to study the potential consequences of the large but geographically uneven decrease in transportation costs generated by containerization.

We imagine a model with many locations, where firms produce a variety of goods under conditions of monopolistic competition and consumers consume a variety of goods and housing. It is costly to move goods between locations, and transportation costs are of the iceberg form. Workers are mobile and move in response to differences in real wages, which depend on a combination of nominal wages, prices for the variety of goods, and land rents.

Given this framework, we ask how the advent of containerization impacts equilib-

[4] It is therefore no surprise that Scottish whiskey bound for US markets was on the first international container trip (Levinson, 2008, p. 165).

[5] We (obviously) haven't written a model for this draft, but plan to for the next draft.

rium population, employment, and wages. Specifically, we assume that some locations—for the moment, an exogenous set of locations—adopt container technology. For locations near containerized ports, the cost of goods transportation declines. Containerization introduces a transportation cost advantage that is greatest at the port and decays with distance to the port.

This reduction in goods transport costs increases the attractiveness of locations near ports for both producers and consumers. For firms, lower trade costs allow producers to "become closer" to their markets, increasing firm market access. For consumers, lower trade costs increase consumers' access to tradeable goods, increasing consumer market access.

Containerization's positive shock to market access causes both an increase in nominal wages and an increase in population and employment at locations near containerized ports. Land rents increase until real wages equalize across locations. As the increase in market access is greatest near containerized ports, the relative gain in population, employment, and nominal wages dissipates as distance to the containerized port increases.

4 Data

To study the impact of this drop in transportation costs on local economic growth, we construct a county-level panel dataset that includes population and employment information, as well as proximity to port and port characteristics. This section gives an overview of the data, and we present full details in the data appendix.

Our sample frame is the county level Decennial Census, for the years 1910 to 2010.[6] We assemble a time invariant panel of counties by aggregating 1950 counties to their 2010 counterparts and by dropping very few counties with large land area changes. From

[6] For the 2010 sample, we use the Decennial Census for population figures and the American Community Survey (years 2008–2012) for other demographic covariates.

1950 to 2010 we observe population, income, and demographic characteristics. We also observe employment, payroll, and employment and payroll by industry from the County Business Patterns. For these data, our time series begins in 1956 and extends annually to 2011.[7] We omit Alaska from our analysis because its administrative districts in 1950 do not correspond to its modern counties. This yields 3,023 counties with complete data.[8]

To this sample frame, we add port attribute data. Our universe of ports is all ports that existed in either 1953 or 2015, as defined by the 1953 and 2015 *World Port Index*. For each port, we observe its location (latitude and longitude), size (in 4 discrete categories), and depth (in 8 discrete categories). We use year of first containerization from the *Containerisation International Yearbook*, volumes 1968 and 1970 to 2010.[9] We also observe 1948 and 1955 international trade in dollars by port from the Census Bureau's Foreign Trade Statistics. We associate each county with a vector of ports and port characteristics, which include the distance from each county to each port, the number of nearby 1953 ports, the maximum 1953 depth of nearby ports, and the values of 1948 and 1955 international trade at nearby ports.[10]

We also include variables that characterize the state of the transportation network now and at the advent of containerization (c. 1957 for highway and c. 1960 for rail). We measure total rail kilometers, highway kilometers, and waterway kilometers in each county.

In addition to these detailed US data, we constructed a less detailed panel dataset of world cities. The sample frame for world cities is the United Nation's 2014 Revision

[7]We received digitized 1956 County Business Patterns data courtesy of Matt Turner and Gilles Duranton. See the data appendix for more information about these data.

[8]Estimations using County Business Patterns data use a slightly smaller sample because the provider suppresses data for counties with very small populations.

[9]For the purposes of this paper, and consistent with the industry definition, we call a port "containerized" when it has special infrastructure and equipment to handle containers. Specifically, the port has invested in equipment to handle shipping containers which enables their movement in and out of ship and onto a train or a truck.

[10]We calculate all distances from the county centroid.

of World Urbanization Prospects. This dataset contains all 1,692 urban agglomerations with populations exceeding 300,000 at any time between 1950 and 2014. By construction, this sample over-represents fast growing cities that were small in 1950 but grew rapidly in the second half of the twentieth century. To mitigate this sampling issue, we restrict the sample to cities with population over 50,000 in 1950, yielding a world panel of 1,051 cities.

5 Empirical Methods

In this section, we explain our empirical strategy for estimating the causal effect of containerization on local economic growth. We first present an OLS framework to analyze the impact of proximity to a containerized port on economic activity to illustrate the strength of the first difference specification. We then discuss potential remaining concerns with causality, followed by a motivation for and details about our instrumental variable strategy.

5.1 First Difference Specification

Our goal is to understand how economic activity responds to the advent of containerization. Empirically, we ask whether county proximity to a containerized port is associated with changes in key economic outcomes, conditional on a host of covariates. Specifically, we estimate

$$\Delta \ln(y_{i,t}) = \beta_0 + \beta_1 \Delta C_{i,t} + \beta_2 X_i + \Delta \epsilon_{i,t} \ , \tag{1}$$

where $i \in I$ indexes counties and $t \in T$ indexes years. Depending on the context, our dependent variable, $y_{i,t}$, will be population, employment, earnings, or other county-level outcomes. The operator Δ denotes long run differences, so that $\Delta \ln(y_{i,t}) = \ln(y_{i,t}) -$

$\ln(y_{i,1950})$.[11] Capital letters denote vectors.

Our key explanatory variable is an indicator for proximity to a containerized port at time t, $\Delta C_{i,t}$, which is equivalent to $C_{i,t}$, as no containerized ports existed in 1950 ($C_{i,1950} = 0 \ \forall i \in I$). We allow for potential non-linear impacts of proximity to a containerized port by creating indicator variables for port proximity by distance bin category. Figure 3a is a graphical representation of this parameterization: counties in the darkest blue are within 100 km of a containerized port, mid-blue are counties from 100 to 200 km of a containerized port, light blue are counties 200 to 300 km of a containerized port, and light red are counties more than 300 km from a containerized port.

Mathematically, we parameterize proximity to a containerized port as

$$\beta_1 \Delta C_{i,t} \equiv \sum_{d \in D} \beta_{1,d} \mathbb{1}\{\text{Closest containerized port is between } d_1 \text{ and } d_2 \text{ km}\}_{i,t}, \qquad (2)$$

where $d \in D$ are a set of distance bins of $\{0-100, 100-200, 200-300\}$ kilometers. We discuss this choice of parameterization in greater detail in the results section. We interpret $\beta_{1,\{0-100\}}$ as the percentage change in the dependent variable in counties within 100 km of a containerized port, conditional on covariates. Coefficients $\beta_{1,\{100-200\}}$ and $\beta_{1,\{200-300\}}$ report coefficients for the other distance bins.

To establish the causal effects of containerization on local economic growth, we must contend with the non-random assignment of containerized ports to counties. This first difference specification nets out any time-invariant county-specific characteristics that are correlated with the location of containerized ports. Such characteristics include geography, proximity to population centers, climate, and historical antecedents for the location of particular industries. This method also nets out any national changes that impact all counties equally from 1950 to 2010.

[11] When we use County Business Patterns data, the initial year is 1956.

In the event that county proximity to a containerized port is also a function of time-varying county attributes, we also include a vector of baseline covariates, X_i. Including initial covariates in the first difference model is akin to allowing for differential growth trends in the dependent variable by the initial covariates. We list these in greater detail in Section 6, but X_i includes regional fixed effects, distance to the ocean, measures of geographic proxity to ports in 1953, the extent of the initial transportation network, initial demographic characteristics, initial industry mix, and pre-1950 county population. We cluster standard errors at the county level, which is equivalent to robust standard errors in this two-period case.

This empirical strategy yields a causal estimate for the effect of proximity to a containerized port on local economic growth when proximity to a containerized port is uncorrelated with the error term. This is equivalent to saying that β_1 can be interpreted as a causal estimate when proximity to a containerized port is randomly assigned, conditional on time-invariant county-level factors and the included initial covariates.

However, suppose that counties specializing in industries with greater trade growth before containerization tend to be closer to containerized ports. This is something we may fail to capture, even after netting out county-specific time-invariant factors and controlling for the initial industry mix. Such endogeneity would yield positive bias in the OLS estimates. Conversely, if counties with worse trade trajectories were more likely to be near a containerized port, OLS estimates would be biased downward.

5.2 Instrumental Variables

To address any remaining non-randomness in the assignment of containerized ports to counties, we use proximity to a very deep port in 1953, Z_i, as an instrument for proximity to a containerized port, $\Delta C_{i,t}$. Specifically, our first stage is

$$\Delta C_{i,t} = \alpha_0 + \alpha_1 Z_i + \alpha_2 X_i + \Delta \eta_{i,t} \ , \tag{3}$$

where $\alpha_1 Z_i$ is

$$\alpha_1 Z_i \equiv \sum_{d \in D} \alpha_{1,d} \mathbb{1}\{\text{Closest very deep port in 1953 is between } d_1 \text{ and } d_2 \text{ km}\}_i. \tag{4}$$

In words, we have three potentially endogenous variables—indicators for counties being within 300 km of a containerized port at time t, in 100 km bins—and three instruments—indicators for counties being within 300 km of a very deep port in 1953, also measured in 100 km bins.

There are two requirements for the instrument to yield a causal estimate of proximity to a containerized port on local economic growth. The first is a strong relationship between proximity to a containerized port and proximity to a very deep port in 1953. The second requirement is that, conditional on covariates, proximity to a very deep port in 1953 is uncorrelated with unobserved determinants of changes in economic activity from 1950 to period t. In other words, proximity to a very deep port in 1953 impacts changes in economic activity only through its impact on proximity to a containerized port, conditional on covariates ($\text{Cov}(Z_i, \Delta\epsilon_{i,t}) = 0$). We discuss each of these requirements in turn.

First, we anticipate that proximity to a containerized port should be strongly related to proximity to a very deep port in 1953 because container ships are much larger than their predecessors. These larger ships sit deeper in the water and thus require greater depth to navigate and dock.

It is possible, but quite expensive, for initially shallow ports to compete as containerized ports by drilling or blasting for dredging. Given enough money and sufficiently lax environmental regulation, a harbor can arguably be made arbitrarily deep. However, port depth is malleable only at great cost. Therefore, initially deep ports have a competitive advantage when technology changes to favor very deep ports. This inability of all ports to adjust equally is confirmed by Broeze, who notes that while "ship designers [keep] turning out larger and larger vessels," and "the engineering limits of port construction and channel deepening have by no means been reached[, t]his, however, may not be said of the capacity of all port authorities to carry the cost of such ventures" Broeze (2002, pp. 175–177). Thus, initial port depth is a key component of the cost of converting a breakbulk port into a containerized port.

The intuition that port depth is a key driver of containerization is borne out in practice by containerization's pattern of adoption. Figure 2a shows the likelihood that a county becomes proximate to (within 300 km of) a containerized port over time by the maximal depth of ports within 300 km of the county in 1953.[12] It is immediately clear that proximity to deep ports in 1953 is a strong predictor of proximity to a containerized port at time t. Counties within 300 km of ports with maximal depth greater than 35 feet are always within 300 km of a containerized port by the end of the sample period. Roughly 20 percent of counties within 300 km of ports with maximal depth between 25 and 35 feet are not near a containerized port by the end of the sample period. For counties within 300 km of less deep ports, however, containerization is decidedly not a certainty. Indeed, counties near initially shallow ports—those less than 20 feet deep—are never within 300 km of a containerized port.

An alternative way to view the strength of our instrument is to compare Figures 3a and 3b. The top panel is the map of US counties, where treated counties are blue and deeper blue indicates greater proximity to a containerized port. The bottom panel

[12] We use depth of the wharf in 1953 as our measure of pre-containerization port depth. Results are generally robust to using anchorage and channel depth, which the *World Port Index* also reports.

repeats this map, but re-colors treated counties in green when the instrument correctly predicts treatment, i.e. when a county is within d_1 and d_2 from the nearest containerized port in 2010 and within d_1 to d_2 from the nearest very deep port in 1953. This picture demonstrates that while the instrument does quite poorly in the midwest, it does very well on the east and west coasts.

Given this evidence of a strong relationship between the endogenous variables and the instruments, we now turn to the second condition for instrument validity—that proximity to a very deep port in 1953 affects local economic growth only through its impact on proximity to a containerized port.

A key concern with the instrument is that port depth may explain county success even before containerization. This is surely true. For many years ports have been engines of growth. However, we rely not on the full distribution of port depth, but on an indicator variable for being a very deep port pre-containerization. Specifically, we call a port "very deep" when it is more than 30 feet deep in 1953. The depth cut-off is beyond what was generally considered a useful depth in the pre-containerization era.

Before containerization, port depth conveyed some advantage, but it was not particularly useful for a port to be very deep. Given the limited draft of breakbulk ships, greater depth was only useful up to a certain point. This is clear even from how data on port depth was collected. The 1953 *World Port Index*'s deepest category is "40 feet and above," while the deepest category in the 2015 *World Port Index* is "76 feet and over." Thus, intuitively, our instrument measures any additional advantage a county gains in containerization by proximity to a very deep port in 1953, conditional on initial covariates. Note that our first difference specification allows for differential growth trends in the dependent variable by the number of ports in 1953 in 100 km bins and the values of international trade at these ports in 1955, also measured in 100 km bins.

Our claim that depths beyond 30 feet were not particularly advantageous to port success is supported by a number of contemporary commentators. A 1938 monograph argues that "For the ports with which we are dealing, the 30-foot channel at low-water will be taken as the minimum standard in relation to the needs of modern ships" (Sargent, 1938).[13] However, he notes that the cost of making a channel deeper is no small endeavor: "It is a question how far the rest of the world, Europe in particular, is prepared, except in special circumstances, to face the very heavy cost of providing for the needs of the ocean mammoth" (Sargent, 1938, p. 21).

This author's focus on the irrelevance of extreme depth is not unique. Even as late as 1952, F. W. Morgan argues in *Ports and Harbours* that beyond a certain level, depth is not a particularly useful feature of a port:

> The importance for a few ports of maintaining a ruling depth sufficient to admit the largest liners [a draft of 40 feet] emphasizes unduly their importance to the port world. A super-liner which comes into a port every few weeks will, it is true, amplify that port's tonnage figures by half a million tons or so annually. . . . The greater part of world trade by sea and the greater part of the traffic of many ports is concerned with ships of more modest size.
>
> It would certainly be possible to devise a classification of ports by the draught of ship which can be berthed in them. Halifax and Wellington would appear in the first class, and their ability to berth the largest ships is a great asset in wartime. It tells, however, only a little about their normal significance as ports. (p. 15, Morgan (1952))

Thus, pre-containerization, being very deep was not a particularly valuable port attribute.

Our instrument is therefore analogous to a cost shifter instrument often used in the industrial organization literature. Port depth should affect the supply of ports after the advent of containerization, but have no effect on the demand for ports.

In Section 6, we empirically allay concerns that the instrument is correlated with pre-containerization local economic activity. To do so, we examine the correlation between

[13]He goes on to write that in the U.S., a 35-foot draught is becoming standard (p. 21).

pre-containerization factors and the identifying variation in the instrument.

6 Results

With this empirical framework in hand, we now turn to estimates of the impact of proximity to a containerized port on county population growth. In the first subsection, we report summary statistics that motivate and give intuition for our findings, followed by OLS results on population. The second subsection then provides empirical justification for the assumption that the instrument is uncorrelated with the error term, and present our main instrumental variable specification. The third subsection considers extensions to this basic empirical framework. Specifically, we investigate impacts over time, we use employment and earnings as dependent variables, and we analyze the sample of world cities. The final subsection explores heterogeneity in the impact of containerization by initial county characteristics.

6.1 OLS Results

We first test the theoretical prediction that proximity to a containerized port causes population growth after the advent containerization. Table 1 illustrates this comparison and previews the main results. The three leftmost columns report county means by distance to the nearest containerized port; the fourth column reports the means for all observations within 300 km of a containerized port, and the final column reports the mean for all other counties, which we call "never containerized." A county may appear in only one distance-to-closest-containerized-port bin. The number of observations in the "ever" and "never" columns sum to the total sample size.

The figures on log population in the first rows of this table clearly show that counties near containerized ports were larger pre-containerization and that counties closest to containerized ports were largest. From 1910 to 1950—the pre-containerization years—log population in counties near future containerized ports is larger and increases at a faster rate than in counties farther from future containerized ports. In addition, the average population among ever-containerized U.S. counties is larger than among counties never near a containerized port. These differences between counties generate a possible bias in the OLS estimation that we address in the empirical section.

The table also shows some additional differences between counties by proximity to a containerized port. Across census regions, counties near containerized ports are over represented in the Northeast, under represented in the Midwest and West, and about proportionately represented in the South. On average, counties near containerized ports had a substantially larger share of workers in manufacturing in 1956. In the final subsection of Section 6 we analyze how county population growth responds to containerization as a function of the initial share of workers in manufacturing.

This table also illustrates our main finding, showing that counties near containerized ports grow at a faster pace after the advent containerization than the average untreated county. This relative increase is also visible in the employment and payroll data from the County Business Patterns data.

Moving to a regression framework, Table 2 presents OLS results from estimating Equation 1. This is a more formal test of the prediction that proximity to a containerized port causes population growth after the advent containerization. Column 1 presents results with controls for regional fixed effects only, and shows a 60 percentage point increase in population for counties within 100 km of a containerized port. This coefficient declines to 35 percentage points for counties within 100 to 200 km of a containerized port and to 24 percentage points for counties 200 to 300 km from a containerized port.

The remaining columns in this table add additional covariates. To address the concern that counties of different size may grow at different rates—concerning because counties near containerized ports are uniformly initially larger—Column 2 controls for log population in years 1920, 1930 and 1940. We also add controls for the share of people with a college degree and share African American, both measured as of 1950. The addition of these controls decreases the estimates of proximity to a containerized port on population growth by less than 15 percent for counties very close (less than 100 km) to containerized ports, and has a negligible impact on counties farther from containerized ports. This confirms the fact that the initial market size in the port hinterland is a determinant of containerization adoption, as we discussed in section 2.

To isolate the impact of containerization from proximity to the coast, initial port intensity, pre-containerization port prominence, and the extent of pre-existing transportation networks, Column 3 adds additional controls. Specifically, these include distance to the ocean, three variables for the number ports in 1953 within 300 km, measured in bins of 100 km, three variables for the total value of 1955 international trade at ports within 300 km, again measured in bins of 100 km, and measures of c. 1950 transportation: the length of highways, navigable waterways, and railways by county. Results decline by about one-third to one quarter, so that the gradient by distance bin is now 44, 25 and 15 percentage points, respectively.

Finally, as we saw in Table 1, counties near future containerized ports had, on average, much higher shares of manufacturing employment in 1956. The final column includes this variable as a control. This addition has little incremental impact on the size of the coefficients. We now estimate 41, 23 and 12 percentage point increases in population with distance to the closest containerized port.

These estimates of population increase are large. The average county grew by 37 percentage points from 1950 to 2010, so these estimates are as large as the average change for counties close to containerized ports, and about one-third of the average increase even for counties within 200 to 300 kilometers of a containerized port.

Consistent with the predictions of the model, we find weaker effects as distance from the containerized port increases. The pattern of decline is monotonic, but not linear. The coefficient for distances 100 to 200 km from a containerized port is slightly more than half of the coefficient for counties within 100 km of a containerized port. The coefficient for the last category, for counties 200 to 300 km from a containerized port, is about half of the previous coefficient. This non-linear pattern of decline suggests that transportation costs increase more than linearly as distance from the port increases. (We return in the discussion of IV results to an explanation of why we choose the 300 km border.)

6.2 IV Results

Although the first difference specification is able to address many potential confounding factors that could be correlated both with proximity to a containerized port and population growth—e.g. past population, initial economic conditions—we believe that there may still be elements in the error term that could be correlated with the treatment variables. For this reason, we now turn to the instrumental variables approach. We start with the graphical reduced form intuition, proceed to instrument strength and validity, and follow with IV results.

To give intuition for the IV analysis, Figure 2b presents a graphical illustration of the "reduced form" regression (a regression of population growth on the instrument). This figure presents the average log population over time by initial depth category. Thick lines indicate counties within 300 km of ports that we classify as very deep in 1953; thin lines are counties within 300 km of ports less than 30 feet deep in 1953 (to be consistent with Figure 2a, we omit counties not near ports in 1953). In essence, the estimation asks whether the thicker lines trend upward more after 1956 (the vertical red line) than the thin lines. This picture shows that they do, and that the gains are driven primarily by initially smaller counties—the beige and purple lines.

We already saw from Figure 2a (discussed in Section 5) that the instrument is strong. Appendix Table 1 validates this intuition, reporting coefficients for the three equations that estimate the full first stage; there is one equation per distance to the closest containerized port. The table shows the pattern we expect if the instrument is working as we hypothesize: counties that are between d_1 to d_2 km from the closest very deep port in 1953 are more likely to be between d_1 to d_2 km from the closest containerized port in 2010. These coefficients on the diagonal are large—in the 0.5 to 0.6 range—and strongly significant. This tells us that, even conditional on the many covariates we use, proximity to a very deep port in 1953 remains an important predictor of proximity to a containerized port in 2010. The lowest F statistic on the instruments in any of these three equations is 72; the highest is 160. Tables with 2SLS estimates always report the Kleinberg-Paap F statistic, which summarizes the overall strength of the first-stage, as suggested by Sanderson and Windmeijer (2016). In our main IV estimates, this F statistic is never less than 70.

Given that the instrument is strong, we now turn to its validity. Specifically, we propose a test to evaluate whether the instruments are correlated with some county-level characteristics that might plausibly be left in the error term. While we cannot do this for all potential confounders, we can observe whether the identifying variation—the residual from a regression of an instrumental variable on the full set of covariates from Table 2—is correlated with specific pre-treatment covariates, also conditional on covariates. Figure 4 uses the instrument $\mathbb{1}\{\text{Closest very deep port in 1953 is between } 0 \text{ and } 100 \text{ km}\}_i$.

Recall that our regression specification controls for log population in 1920, 1930, and 1940. Were the identifying variation in the instrument to be related to the log of 1910 population, this would suggest that the pre-treatment controls were not adequately capturing the historical pattern of population growth. We do not find this to be the case. Figure 4a shows the identifying variation from the instrument on the y axis, and the residual from a regression of the log of 1910 population on covariates on the horizontal axis. We find no significant relationship ($t = 0.00000005$) between these two variables.

Similarly, recall that the regression controls for the 1955 value of international trade flows in each of the three distance-to-containerized-port bins. If this covariate did not sufficiently control for the impact of pre-containerization port strength on population growth, we would expect that the identifying variation would be related to the 1948 value of international trade flows by distance-to-containerized-port bins. Figure 4b shows that this is not the case. Here, the relationship between the identifying variation and the value of 1948 trade within 100 km of a containerized port, conditional of covariates, is insignificant ($t = -0.00000003$).

In fact, we do all six estimations like this: a regression of the identifying variation from each of the three instruments (one per distance bin) on 1910 population and on the dollar value of 1948 international trade flows at ports in the same distance interval as the instrument, conditional on covariates. In these six regressions, we find no statistically significant relationship (the largest t-statistic is equal to $t = -0.00000001$; see all plots in Appendix Figure 1).

We report IV results in the second half of Table 2. The columns repeat the pattern of covariates from the first half of the table. The coefficients are generally quite similar, though slightly larger. Why are the IV results larger than the OLS results? Suppose that larger counties grow at a slower rate than smaller counties and that larger counties are more likely to be near a containerized port. When we correct for this endogeneity with the instrument—in principle, giving larger weight to smaller counties, where the depth is the main driver of the containerization decision—the coefficient should increase. The most fully saturated model shows a 47 percentage point increase in population over the 60 years from 1950 to 2010 for counties within 100 km of a containerized port. This coefficient declines to 29 and 20 percentage points, as we expect, for counties slightly farther from containerized ports.[14]

Even with the instrument, some potential concerns with these estimates remain. We use Table 3 to alleviate some concerns with the IV estimates. Rappaport and Sachs (2000) find that from 1960 to 2000, population gains are larger in coastal locations. Our estimation deals with this threat though the first difference specification, which nets out any county-specific time-invariant impact of a coastal location. In addition, our main specification allows for population growth to differ by distance to the ocean and initial port intensity.

To further isolate the impact of containerization from proximity to the coast, Table 3's column 2 restricts the sample to counties within 400 kilometers of a port in 1953. The sample size drops from 3,023 to 1,767 observations, but the coefficients decline only slightly (compare estimates to column 1, which repeats the most complete specification from Table 2). This suggests that population growth in counties near a containerized port is not driven by a comparison with slower-growing centrally located counties.

Alternatively, we know from the summary statistics in Table 1 that counties near containerized ports experience more rapid population growth pre-containerization, and this trend may have continued after 1956 irrespective of containerization. We account for this in the main estimates by including log population in 1920, 1930, and 1940. In Table 3, column 3 additionally includes squares of those measures of past population, in the event that previous population impacts population growth non-linearly. Again, the estimates are little changed.

[14]Both here and in the OLS estimates, we compare counties within 300 km of a containerized port to all other counties. As theory does not provide guidance on the physical distance over which containerization might have a measurable impact, we turn to the data as a guide. Appendix Figure 2 shows regression coefficients from a version of Equation 1 where distance to containerized port is measured in 50 km bins. Gray bands are confidence intervals. These results show that the association between proximity to a containerized port and population growth is indistinguishable from zero at 300 km. In our main specification, we use bins of 100 km, rather than the smaller 50 km ones, to increase the power in the estimates. This is particularly important when we examine whether containerization's impacts differ by initial conditions in subsection 6.4.

A potential concern with our instrument is its failure to predict containerization in the Great Lakes midwest. Shallow ports there do containerize, and the region experiences the slowest population growth over our period of analysis. To allay fears that the results are driven by this potentially anomalous treatment of the midwest, column 4 omits the midwest region entirely, leaving 1,975 observations. Results in this column are smaller than the original specification, but the pattern of decline with distance to the closest containerized port remains. Indeed, we should expect smaller coefficients in this estimation because the control group—non-midwest, non-containerized ports—now has a higher average population growth. Note the increase in the mean of the dependent variable from 0.373 to 0.508 (final row of the table). Still, we observe a relative population increase of 37 percentage points near containerized ports, an increase of almost three-quarters of the mean.

Recent work in urban economics strongly suggests that growth is associated with an area's education and demographic characteristics (Glaeser and Saiz, 2004; Florida, 2002; Moretti, 2004). Column 5 includes additional controls for the share of people 25 or older with a high school degree, the share foreign born, the number of government workers per capita, and the share age 65 and older. The addition of these covariates decreases the coefficients slightly, with greater impact for the category closest to containerized ports. The coefficients remain sizeable, and retain the pattern of decline with distance to containerized ports.

6.3 Extensions

Our results show that proximity to a containerized port causes population growth. We now test the theoretical predictions that proximity to a containerized port causes an increase in employment (which is equivalent to population in the model) and an increase in nominal wages. In the remainder of this subsection, examine the impact of containerization on income, analyze the impact of containerization over time, and use an alternative data source to examine the impact of containerization on city growth across the globe.

We begin by testing whether proximity to a containerized port is associated with an increase in employment in Table 4. The first two columns in this table use data from the County Business Patterns. These data measure employment and payroll by county and industry.

We confirm that employment increases from 1956 to 2011 in counties near containerized ports in Column 1, using the IV estimation with the full set of covariates from Table 2. While only the first coefficient is statistically significant, the magnitude and pattern of employment increases is strikingly similar to what we find from Census data. However, the average percentage point increase in employment over the period is substantially larger, at 1.13 (see final row), than the average increase in population (0.37, see final row in Table 3).

The dependent variable in Column 2 is nominal payroll per employee. In addition to increases in employment, proximity to a containerized port also causes an increase in nominal payroll per employee. Counties closest to containerized ports see increases of 12 percentage points, relative to a mean of 240 percent. The remaining two distance categories see increases of 7 and 4 percentage points; only the former is statistically significant.

To corroborate these impacts on this proxy for wages, we return to data from the Decennial Census. Using binned income data, we estimate the 10th, 50th and 90th percentiles of the income distribution in each county (see Data Appendix for further details). For all percentiles of the income distribution, income increased more, on average, than payroll per employee (see final row). Proximity to a containerized port is causes an increase in income at all points of the income distribution, though the coefficients are larger at the 10th and 50th percentiles. Like all our other outcomes, there is a decline in impact with distance to the nearest containerized port.

The model also suggests that changes in population and employment should be increasing over time, as workers migrate to higher wage areas. Furthermore, the impact of containerization should depend on the extent of the container network—the value of proximity to a containerized port increases with the number of destinations that can handle containers across the globe. To test for these changes in the impact of containerization, we re-estimate Equation 1, using different final years, starting in 1970. We report coefficients from these estimations in Figure 5, which displays results by decade (each vertical line is a separate estimation). Full circles are significant coeffcients (at the five percent level); hollow circles are insignificant coefficients. The red line at the top reports the coefficients for counties from 0 to 100 km of a containerized port; the orange line 100 to 200, and the yellow line 200 to 300 km.

Apart from a blip in 1980, counties near containerized ports have large population gains that increase over time. For example, in 1970, only 15 years after the advent of containerization, counties closest to containerized ports had grown by about 25 additional percentage points. By 2010, this figure was 45. While estimates for counties farther from ports are smaller, they also follow this general pattern of increase.

In this paper, we focus primarily on the United States, due to the rich data available at a relatively small geographic scale. However, containerization is clearly a global phenomenon, and one that may have even more power over economic activity in countries other than the United States. To give some baseline of containerization's impact on city growth across the globe, we use world population data to estimate regressions that are parallel to the US ones.

We report results in Table 5. The first column reports OLS results without covariates, and shows population declines from 1950 to 2010 in cities near a containerized port using the same distance bins as the US sample. The second column adds the same controls as in the US sample for the number of ports in 1953 near each city, in three 100 km distance bins. This increases the coefficient somewhat, suggesting that non-port adjacent cities grew at a faster rate than port-adjacent cities, on average.

The third column adds country dummies. This increases the coefficients at all distance bins, suggesting that population growth from 1950 to 2010 was faster, on average, in countries with fewer containerized port cities. Within country, however, proximity to containerized ports causes more population growth. The final column allows for differential growth by initial population. In this specification, cities within 100 km of a containerized port increase in population.

Just as in the US sample, we are concerned that the assignment of containerized ports to cities is not random, generating bias in the OLS results. We use the same instrumenting technique that we use for the US sample[15], and find, like we do for the US, that proximity to a very deep port in 1953 is strongly related to proximity to a containerized port in 2010 (see Appendix Figure 3; Appendix Table 3 presents summary statistics and Appendix Table 4 shows a strong first stage).

Using this instrumental variable strategy produces results with the same signs as the OLS ones, but substantially larger in magnitudes. In the most complete specification, we find that cities within 100 km of a containerized port grow by an additional 25 percentage points, or about 16 percent of the mean. For cities within 100 to 200 km and within 200 to 300 km of a containerized port, we estimate a population growth increase of 23 percentage points. This finding is smaller in absolute terms than the US one, and is much smaller than relative to the average growth in world cities, which is about 150 percentage points over the sixty year period.

[15]We actually haven't cleaned up the port data sufficiently to make this entirely parallel; this is a job for the next revision.

6.4 Impact of Containerization by Initial Conditions

In the previous subsections, we have shown that proximity to a containerized port causes population, employment, and wages to increase. These estimates are averages across treated counties. However, the distribution of port traffic is quite uneven, with a small number of large ports making up the lion's share of volume. In this section, we consider whether containerization-induced economic growth is spread evenly across treated counties, or whether it is concentrated in counties with particular initial conditions.

To do this, we consider variables $x_i \in X_i$, which we observe circa 1950. For example, x_i may be the share of workers in manufacturing in the county in 1956. We create an indicator variable, h_i, which takes on the value 1 when a county's share of workers in manufacturing is less than the median share across all counties within 300 km of a containerized port (or $h_i = 1$ when $x_i < \text{median}(x_i)$).[16] We then include this indicator variable h_i and its interaction with the variables of interest, $h_i * \Delta C_{i,t}$, in Equation 1.[17]

We report the interacted and uninteracted coefficients on the measure of proximity to a containerized port ($\Delta C_{i,t}$) in Table 6, where the dependent variable is again the log population. In the first column, we examine heterogeneity by the initial strength of the manufacturing sector. We find that half of the containerization-induced population growth in counties within 100 km of containerized ports occurs in counties with lower than median share of workers in manufacturing. For counties slightly farther from the port, almost all of the containerization-induced population growth occur in counties with lower than median share of workers in manufacturing in 1956.

While no initial condition explains as much of containerization-induced population growth as an initially small manufacturing sector, we do observe some differen-

[16]We define the median among only treated counties. We do this to be consistent, because the median of some variables (for example, international trade at nearby ports in 1955) is not defined for untreated counties. Results are largely similar when we use the median across all counties.

[17]We do a similar interaction for the instrument, so that the instruments are now $h_i * Z_i$ and Z_i.

tial growth by other initial conditions. In general, containerization-induced population growth is somewhat more concentrated in counties with initially smaller populations, counties with lower international trade at nearby ports in 1955, counties with initially fewer kilometers of highways, and counties with a more educated population.

Overall, this table paints a picture of containerization exerting the greatest influence not in dominant agglomerations—large, wealthy urban areas—but in second-tier agglomerations. These second-tier agglomerations are initially smaller and less concentrated in the vanguard technology of the 1950s (manufacturing), but have a more educated population. One hypothesis that rationalizes these results (that we hope to test in future drafts) is that this pattern is generated by containerization's demand for large areas of land. Thus, containerization can only be implemented well in areas with initially low land values.

To explore whether containerization-induced gains to income are similarly concentrated in these second tier agglomerations, Table 7 focuses on the two initial conditions from the previous table most correlated with containerization-induced gains in population: the initial share of manufacturing employment and the initial share of people with less than a high school education. The first three columns of the table test whether income increases—at the 10th, 50th and 90th percentiles—are disproportionately in counties with initially lower than median share of manufacturing employment.

Containerization-induced income gains in Table 7 are not as unevenly distributed with initial share of manufacturing employment as were the containerization-induced population gains in Table 6. While all the interaction coefficients are positive—suggesting at least some greater growth in counties with initially weaker manufacturing sectors—their magnitude is generally small relative to the overall coefficients.

In contrast, containerization-induced income gains are concentrated in counties where the share of people with less than a high school degree is lower than the median, or where the population is more educated. In the final three columns of Table 7, the interacted coefficients are uniformly larger than the uninteracted coefficients. In some cases, the interacted coefficients are more than twice as large as the uninteracted coefficients. While all containerized counties and all points in the income distribution see containerized-induced gains in income, these gains are concentrated in initially more educated counties.

This concentration is most marked in the 10th and 50th percentiles. The final row of Table 4 reports mean increases in nominal income by income percentile; average gains are largest for the tenth percentile over this period. Containerization-induced gains in 10th percentile income for counties with more than the median high school education share are about ten percent of overall nominal income gains.

7 Conclusion

We analyze how local economic growth responds to the dramatic change in transportation cost wrought by containerization. We use a novel cost-shifter instrument to show that, consistent with the predictions of a standard new economic geography model, counties near containerized ports have large increases in population and employment, and a somewhat smaller positive impact on wages. These population gains are located predominantly in counties that were initially less manufacturing intensive. Income gains are concentrated in counties with initially higher levels of education.

References

Allen, Treb and Arkolakis, Costas, 2014. "Trade and the Topography of the Spatial Economy." *Quarterly Journal of Economics* 129(3): 1085–1140.

Autor, David H., Dorn, David, and Hanson, Gordon H., 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review* 103(6): 2121–68.

Baum-Snow, Nathaniel, 2007. "Did Highways Cause Suburbanization?" *Quarterly Journal of Economics* 122(2): 775–805.

Bernhofen, Daniel, El-Sahli, Zouheir, and Kneller, Richard, 2016. "Estimating the Effects of the Container Revolution on World Trade." *Journal of International Economics* 98: 36–50.

Bridgman, Benjamin, 2014. "Why Containerization Did Not Reduce Ocean Trade Shipping Costs." Working Paper.

Broeze, Frank, 2002. *The Globalisation of the Oceans: Containerisation From the 1950s to the Present*, vol. 23 of *Research in Maritime History*. St. John's, Newfoundland: International Maritime Economic History Association.

Dix-Carneiro, Rafael and Kovak, Brian, 2015. "Trade Reform and Regional Dynamics: Evidence From 25 Years of Brazilian Matched Employer-Employee Data." Working paper.

Donaldson, Dave, forthcoming. "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure." *American Economic Review* .

Donaldson, Dave and Hornbeck, Richard, 2016. "Railroads and American Economic Growth: A Market Access Approach." *Quarterly Journal of Economics* .

Duranton, Gilles, Morrow, Peter M., and Turner, Matthew A., 2014. "Roads and Trade: Evidence from the US." *Review of Economic Studies* 81: 681–724.

Duranton, Gilles and Turner, Matthew, 2012. "Urban Growth and Transportation." *Review of Economic Studies* 79: 1407–1440.

Florida, Richard, 2002. *The Rise of the Creative Class*. New York, NY: Basic Books.

Fujita, Masahisa, Krugman, Paul, and Venables, Anthony J., 1999. *The Spatial Economy: "Cities, Regions and International Trade"*. The MIT Press.

Glaeser, Edward L. and Saiz, Albert, 2004. "The Rise of the Skilled City." *Brookings-Wharton Papers on Urban Affairs* pages 47–105.

Helpman, E., 1995. "The Size of Regions." Tech. rep.

Hummels, David, 2007. "Transportation Costs and International Trade in the Second Era of Globalization." *Journal of Economic Perspectives* 21(3): 131–154.

Hummels, David and Schaur, Georg, 2013. "Time as a Trade Barrier." *American Economic Review* 103(7): 2935–2959.

Krugman, Paul, 1991. "Increasing Returns and Economic Geography." *Journal of Political Economy* 99(3): 483–99.

Levinson, Mark, 2008. *The Box: How the Shipping Container Made the World Smaller and the World Economy Bigger*. Princeton, NJ: Princeton University Press.

Moretti, Enrico, 2004. "Human Capital Externalities in Cities." In Vernon Henderson and Jacques Thisse, (Eds.) "Handbook of Regional and Urban Economics," Elsevier.

Morgan, F. W., 1952. *Ports and Harbours*. London, UK: Hutchison's University Library.

National Geospatial Intelligence Agency, 1953. "World Port Index, 1953." H.o. pub. no. 950, first edition.

National Geospatial Intelligence Agency, 2015. "World Port Index, 2015." Twenty-fourth edition.

Pascali, Luigi, 2014. "The Wind of Change: Maritime Technology, Trade and Economic Development." Working Paper 165, CAGE.

Rappaport, Jordan and Sachs, Jeffrey D., 2000. "The United States as a Coastal Nation." *Journal of Economic Growth* 8: 5–46.

Redding, Stephen and Sturm, Daniel M., 2008. "The Costs of Remoteness: Evidence from German Division and Reunification." *American Economic Review* 98(5): 1766–1797.

Redding, Stephen J., 2016. "Goods trade, factor mobility and welfare." *Journal of International Economics* 101: 148 – 167.

Redding, Stephen J. and Turner, Matthew A., 2015. "Transportation Costs and the Spatial Organization of Economic Activity." In Gilles Duranton and William Strange, (Eds.) "Handbook of Regional and Urban Economics," Elsevier B. V.

Rodrigue, Jean-Paul, 2015.

Rua, Gisela, 2014. "Diffusion of Containerization." Finance and Economics Discussion Series 88, Board of Governors of the Federal Reserve System (U.S.).

Sanderson, Eleanor and Windmeijer, Frank, 2016. "A weak instrument F-test in linear IV models with multiple endogenous variables." *Journal of Econometrics* 190(2): 212–221.

Sargent, A. J., 1938. *Seaports and Hinterlands*. London, UK: Adam and Charles Black.

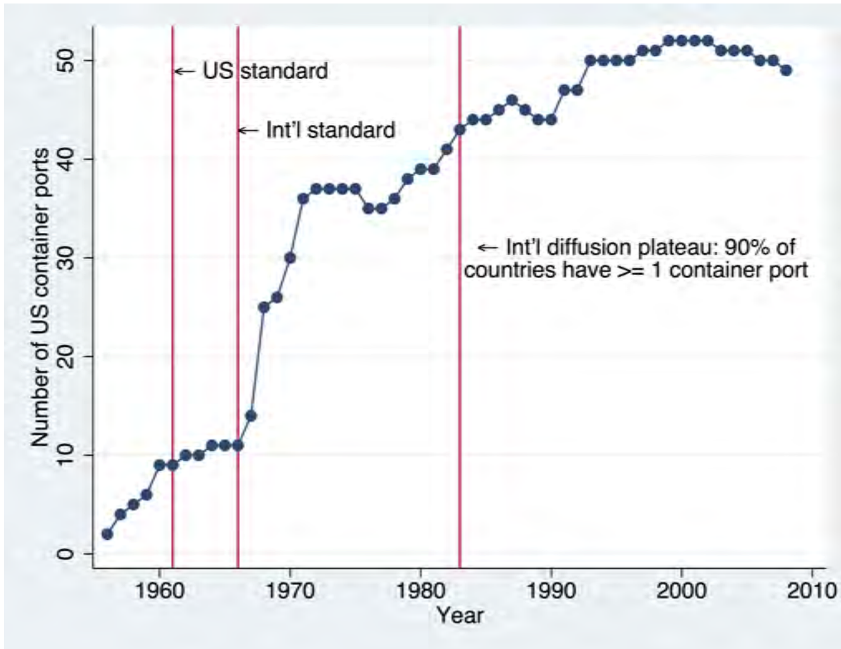
Topalova, Petia, 2010. "Factor Immobility and Regional Impacts of Trade Liberalization: Evidence on Poverty from India." *American Economic Journal: Applied Economics* 2(4): 1–41.

United Nations Conference on Trade and Development, 2013. *Review of Maritime Transport*. New York, NY: United Nations.

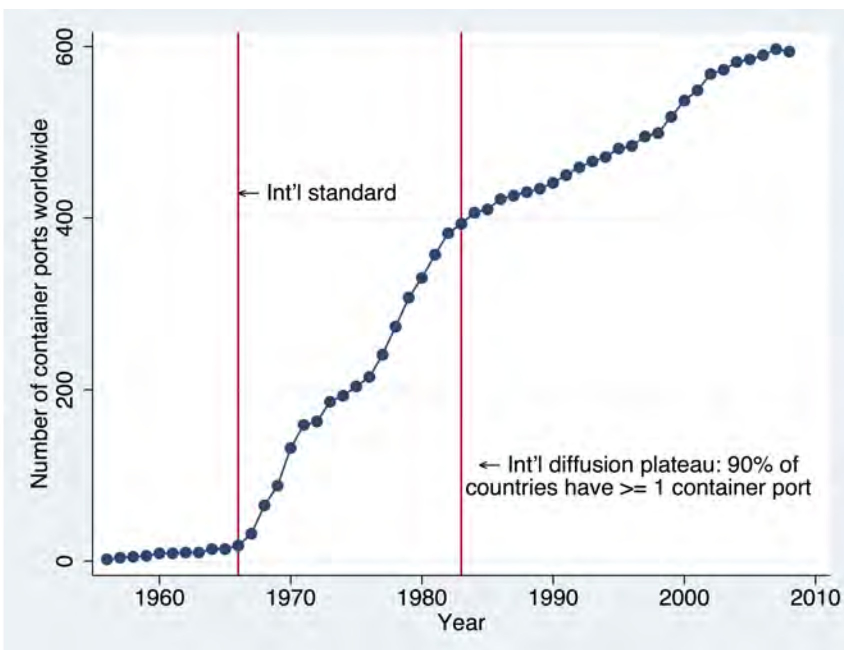
Wilson, R. P., 1982. "The Containerization of Ports." UC Santa Cruz Working Paper.

Figure 1: Adoption of Containerization: 1956–2008

(a) United States



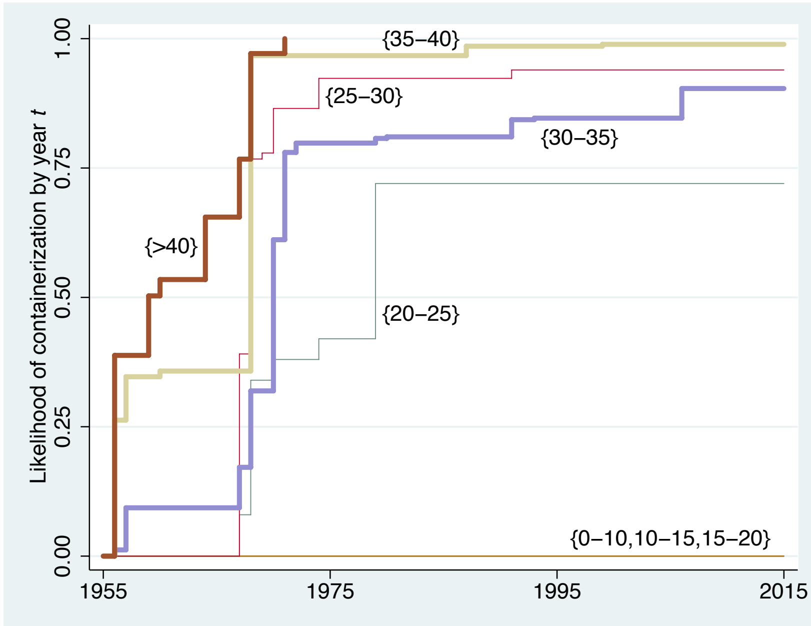
(b) Worldwide



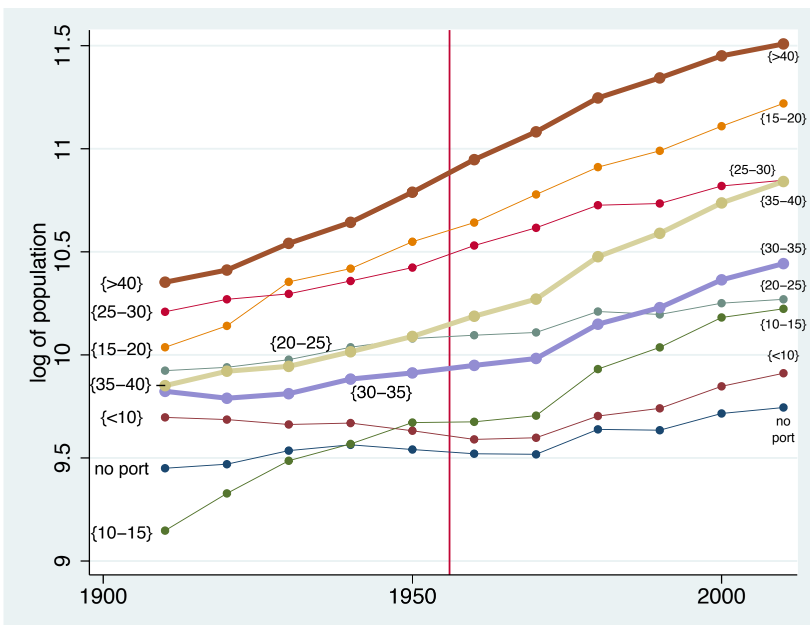
Source: *Containerisation International Yearbook*, volumes 1968 and 1970–2010.

Figure 2: Graphical Intuition

(a) First Stage: Depth and Likelihood of Containerization



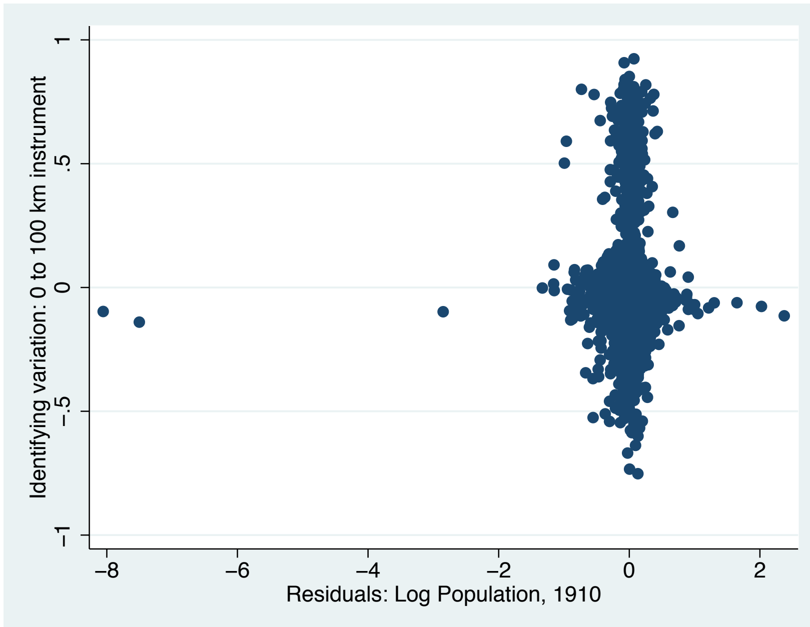
(b) Reduced Form: Depth and Population Changes



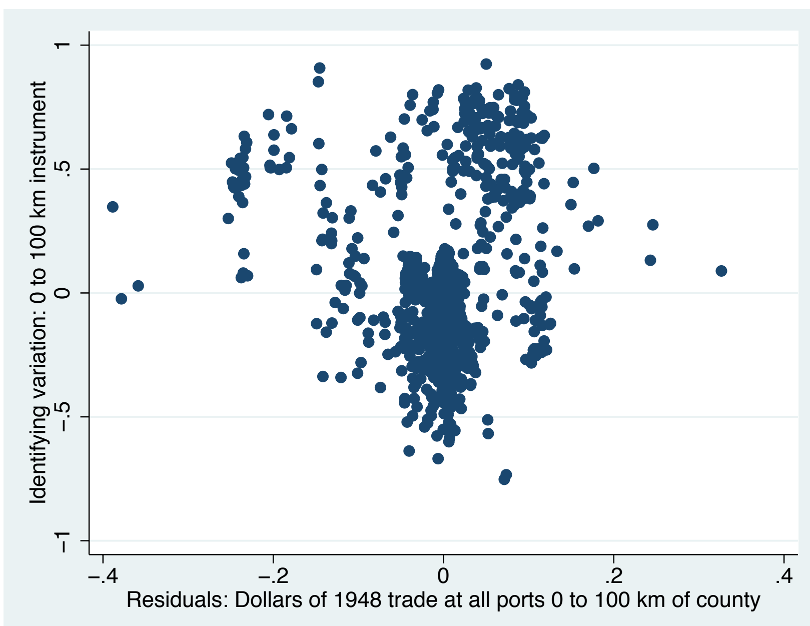
Notes: In both figures, thick lines are depths that we label "very deep" in out estimation. Figure 2a shows the likelihood that a county will have a containerized port within 300 km in year *t* by the 1953 depth of the nearest port within 300 km. On average, deeper ports are more likely to ever containerize, and more likely to containerize early. Figure 2b reports the log population by the 1953 depth of the nearest port within 300 km, as defined in Figure 2a. The ratio of these impacts is the unconditional IV estimate.

Figure 4: Instrument Variation vs. Pre-Treatment Covariates

(a) Versus Log Population, 1910

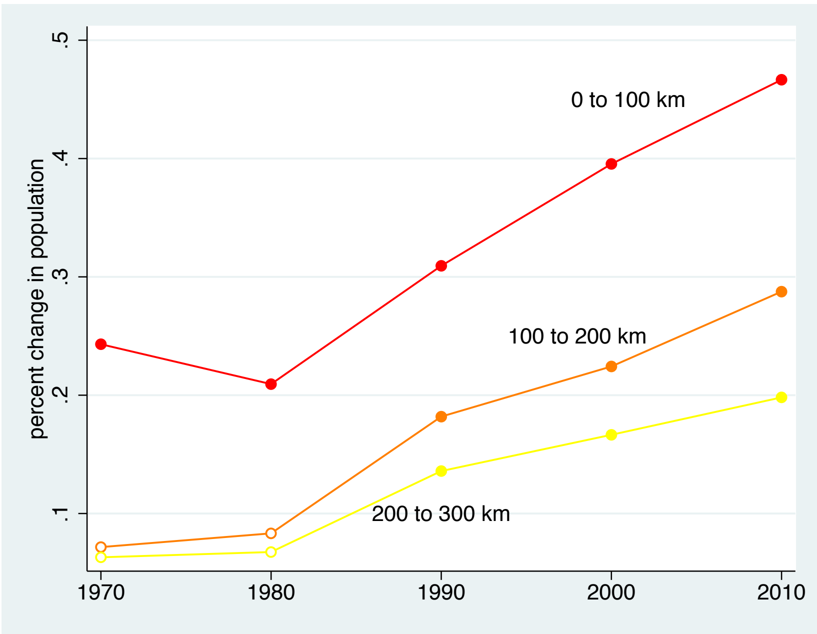


(b) Versus Billions of Dollars of 1948 International Trade at Port within 100 km



Notes: "Identifying variation" is the residual from a regression of the instrument (county is within 0 to 100 kilometers of a "very deep" port in 1953) on the full set of covariates from equation 1 (as in Table 2, columns 4 and 8).

Figure 5: Containerization's Impact Increases Over Time



Notes: This picture reports coefficients from the specification in column 8 of Table 2, but where the dependent variable is the change in log population from 1950 to the year reported on the horizontal axis and the endogenous variable is the change in containerization status from 1950 to the year reported on the horizontal axis. Each dot corresponds to an estimated coefficient for each distance bin. Full circles are significant at the 5 percent level; hollow circles are insignificant coefficients.

Table 1: County Characteristics by Distance to Closest Containerized Port

	Distance to Containerized Port, km				
	0 to 100	100 to 200	200 to 300	Ever Cont.	Never Cont.
	(1)	(2)	(3)	(4)	(5)
Log Population					
1910	10.31	10.03	10.02	10.11	9.47
	[1.22]	[0.82]	[0.80]	[0.95]	[0.96]
1920	10.39	10.05	10.04	10.14	9.49
	[1.31]	[0.88]	[0.85]	[1.02]	[0.93]
1930	10.54	10.11	10.05	10.21	9.56
	[1.38]	[0.90]	[0.92]	[1.08]	[0.88]
1940	10.63	10.17	10.11	10.28	9.60
	[1.40]	[0.91]	[0.91]	[1.09]	[0.90]
1950	10.81	10.23	10.14	10.36	9.58
	[1.47]	[0.97]	[0.97]	[1.16]	[0.96]
2010	11.70	10.75	10.52	10.94	9.79
	[1.50]	[1.16]	[1.15]	[1.35]	[1.32]
Log Employment					
1956	9.02	8.19	8.04	8.37	7.18
	[1.94]	[1.44]	[1.45]	[1.65]	[1.43]
2011	10.37	9.31	9.08	9.53	8.35
	[1.83]	[1.45]	[1.47]	[1.66]	[1.55]
Log First Quarter Payroll, $1000s					
1956	8.75	7.82	7.64	8.02	6.69
	[2.17]	[1.63]	[1.65]	[1.86]	[1.62]
2011	12.55	11.35	11.10	11.61	10.28
	[2.04]	[1.55]	[1.59]	[1.81]	[1.71]
Region					
Northeast	0.19	0.17	0.12	0.16	0.00
Midwest	0.19	0.28	0.38	0.29	0.39
South	0.49	0.48	0.45	0.47	0.43
West	0.13	0.07	0.05	0.08	0.17
Share Employment,	0.43	0.42	0.42	0.42	0.26
Manuf., 1956	[0.19]	[0.19]	[0.20]	[0.19]	[0.22]
Observations	370	523	442	1335	1688

Note: This table reports means, and standard deviations in brackets. The number of observations at the bottom of the table applies to all variables except the 1910 population and the payroll and employment variables; each has slightly fewer observations. Payroll is total first quarter payroll.

Table 2: Containerization Associated with Increased Population, Particularly Near the Port

	OLS				IV			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Closest container port is								
0 to 100 km	0.684***	0.604***	0.440***	0.413***	0.685***	0.642***	0.338***	0.467***
	(0.039)	(0.040)	(0.053)	(0.053)	(0.054)	(0.057)	(0.122)	(0.120)
100 to 200 km	0.348***	0.351***	0.250***	0.233***	0.237***	0.371***	0.221**	0.288***
	(0.033)	(0.032)	(0.041)	(0.041)	(0.056)	(0.055)	(0.088)	(0.087)
200 to 300 km	0.235***	0.202***	0.149***	0.116***	0.215***	0.267***	0.175**	0.198**
	(0.033)	(0.033)	(0.038)	(0.038)	(0.064)	(0.067)	(0.086)	(0.086)
Covariates								
Regional fixed effects	x	x	x	x	x	x	x	x
Log population, 1920-1940		x	x	x		x	x	x
Demographics		x	x	x		x	x	x
Distance to the ocean			x	x			x	x
Number of 1953 ports			x	x			x	x
Total int'l trade at ports, 1955			x	x			x	x
1950s-era transportation			x	x			x	x
Share manufacturing employment, 1956				x				x
R-squared	0.186	0.328	0.353	0.370	0.183	0.327	0.352	0.369
Kleinberg-Paap F Stat					251.2	241.6	73.9	70.2

Notes: All regressions use 3,023 observations. The dependent variable is log population, with a mean of 0.373. Stars denote significance levels: * 0.10, ** 0.05, and *** 0.01. Demographics is share of people with a college degree or more and share African America, both measured as of 1950. Number of 1953 ports and total international trade at ports, 1955 are both vectors, with totals by 100 km bins. 1950s-era transportation is a vector which measures the kilometers of highways, kilometers of navigable waterways, and kilometers of railroads in each county, c. 1950. See data appendix for complete details on years and sources.

Table 3: Impact of Containerization Robust to Alternative Specifications

	Original spec.	Within 400 km of a 1953 port	Squares of population	Omit midwest region	Additional demographic covariates
	(1)	(2)	(3)	(4)	(5)
Closest Container Port is within					
0 to 100 km	0.467***	0.418***	0.387***	0.367***	0.384***
	(0.120)	(0.129)	(0.117)	(0.127)	(0.120)
100 to 200 km	0.288***	0.218**	0.225***	0.170*	0.232***
	(0.087)	(0.098)	(0.087)	(0.097)	(0.087)
200 to 300 km	0.198**	0.199**	0.162*	0.092	0.152*
	(0.086)	(0.098)	(0.085)	(0.087)	(0.085)
R-squared	0.369	0.324	0.383	0.308	0.382
Kleinberg-Paap F Stat	70.2	65.4	69.5	76.9	70.8
Observations	3023	1767	3023	1975	3023
Mean, dependent variable	0.373	0.514	0.373	0.508	0.373

Notes: Stars denote significance levels: * 0.10, ** 0.05, and *** 0.01. These specifications are all IV regressions, using the log population as the dependent variable and including the most complete covariate list from Table 2. Column 1 repeats the most saturated estimation from Column 8 Table 2. Column 2 restricts the sample to counties within 400 km of a 1953 port. Column 3 includes squares of 1920, 1930 and 1940 population. Column 4 omits the midwest census region, which has no very deep ports. Column 5 includes additional demographic covariates measured in 1950: share of people 25 or older with less than a high school degree, share foreign born, government workers per capita, and share age 65 and older.

Table 4: More Employment and Higher Earnings Near Containerized Ports

	IV, Dependent Variable is				
			Log pth percentile income, where p is		
	Log employment	Log payroll/ employee	10	50	90
	(1)	(2)	(3)	(4)	(5)
Closest container port is					
0 to 100 km	0.375**	0.120**	0.228***	0.348***	0.210***
	(0.147)	(0.054)	(0.064)	(0.053)	(0.041)
100 to 200 km	0.168	0.070*	0.123*	0.154***	0.117***
	(0.109)	(0.040)	(0.048)	(0.039)	(0.032)
200 to 300 km	0.051	0.041	0.084+	0.081*	0.016
	(0.104)	(0.038)	(0.048)	(0.040)	(0.033)
R-squared	0.180	0.154	0.280	0.403	0.316
Kleinberg-Paap F Stat	70.9	70.9	70.1	70.1	70.1
Observations	2981	2981	3022	3022	3022
Mean, Dependent Variable	1.133	2.448	3.547	3.147	3.176

Notes: These specifications are all IV regressions including the most complete covariate list from Table 2.Column 1 and 2 use CBP data and look at changes from 1956 to 2011. Column 3, 4, and 5 use census income data and look at changes from 1950 to 2010. Stars denote significance levels: * 0.10, ** 0.05, and *** 0.01.

Table 5: Change in Log Population, 1950 to 2010, by Distance to Containerized Port (World Sample)

	OLS Estimates				IV Estimates			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ever Cont. (0-100km)	-0.0613	0.0362	0.0817*	0.1454***	-0.8494***	-0.5278***	0.1695	0.2488**
	(0.0656)	(0.0617)	(0.0491)	(0.0488)	(0.1473)	(0.1370)	(0.1078)	(0.1037)
Ever Cont. (100-200km)	-0.1910***	-0.0252	0.0436	0.0183	-0.5693***	-0.2839**	0.2938**	0.2343**
	(0.0671)	(0.0637)	(0.0528)	(0.0511)	(0.1481)	(0.1414)	(0.1155)	(0.1084)
Ever Cont. (200-300km)	-0.1553**	-0.0389	0.0929*	0.0790	-1.0758***	-0.6995***	0.2958**	0.2297*
	(0.0661)	(0.0609)	(0.0564)	(0.0529)	(0.1558)	(0.1434)	(0.1355)	(0.1274)
Number of ports in 1953	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Country dummies	No	No	Yes	Yes	No	No	Yes	Yes
Log population in 1950	No	No	No	Yes	No	No	No	Yes

Notes: Sample of world cities over 50K in 1950. Dependent variable: Change in log population between 1950 and 2010. All specifications include a dummy variable for having a port within 300km in 1953. The mean of the dependent variable is 1.54. All regressions have 1051 observations. Robust standard errors in parentheses. Stars denote significance levels: * 0.10 ** 0.05 *** 0.01. Sources: See data appendix.

Table 6: Variation in Impact By 1950 Status

	1950 Interaction Variable is					
	Manuf. share of Employmt	1950 Population	Trade of ports w/i 300 km	Total highway length	Total rail length	Share people < HS degree
	(1)	(2)	(3)	(4)	(5)	(6)
Closest container port is within						
0 to 100 km	0.22	0.227	0.474***	0.073	0.492***	0.539***
	(0.143)	(0.156)	(0.122)	(0.163)	(0.145)	(0.12)
100 to 200 km	-0.063	0.212*	0.279***	0.045	0.365***	0.227***
	(0.101)	(0.121)	(0.088)	(0.159)	(0.116)	(0.086)
200 to 300 km	-0.094	0.138	0.215**	-0.026	0.286**	0.137
	(0.094)	(0.111)	(0.108)	(0.134)	(0.115)	(0.084)
Container port distance * 1{County ≤ median(column variable)}						
0 to 100 km	0.288***	0.395***	0.280*	0.479***	-0.035	-0.009
	(0.108)	(0.124)	(0.154)	(0.118)	(0.102)	(0.138)
100 to 200 km	0.525***	0.112	0.103	0.257*	-0.117	0.526***
	(0.104)	(0.108)	(0.108)	(0.137)	(0.1)	(0.157)
200 to 300 km	0.438***	0.12	0	0.258*	-0.136	0.437***
	(0.125)	(0.127)	(0.105)	(0.139)	(0.124)	(0.165)
R-squared	0.381	0.38	0.365	0.38	0.368	0.358
Kleinberg-Paap F Stat	30.4	37.5	16.7	37	25.4	14.9
Median, interaction variable	0.444	10.2	0.328	0	103	0.725
Share of observations ≤ median						
0 to 100 km	0.483	0.402	0.302	0.664	0.501	0.588
100 to 200 km	0.49	0.514	0.454	0.811	0.495	0.48
200 to 300 km	0.526	0.567	0.725	0.81	0.506	0.449

Note: Trade at ports within 300 km is measured as of 1955. Highway miles measured c. 1960. Rail measured c. 1957. All regressions have 3,023 observations.

Table 7: Changes in Income by Distribution Points and Initial Manufacturing and Income

	Dependent Variable Income Percentile is					
	10	50	90	10	50	90
	1950 Interaction Variable is					
	1956 Manuf. Share of Employment			Share people < HS degree		
	(1)	(2)	(3)	(4)	(5)	(6)
Closest container port is within						
0 to 100 km	0.177**	0.281***	0.177***	0.199***	0.292***	0.158***
	(0.081)	(0.062)	(0.05)	(0.066)	(0.054)	(0.042)
100 to 200 km	0.067	0.089*	0.098**	0.093*	0.130***	0.110***
	(0.065)	(0.051)	(0.039)	(0.048)	(0.039)	(0.032)
200 to 300 km	-0.002	-0.016	-0.026	0.052	0.047	0.002
	(0.062)	(0.05)	(0.039)	(0.05)	(0.041)	(0.031)
Container port distance * 1{County $\leq$ median(column variable)}						
0 to 100 km	0.057	0.077	0.038	0.308***	0.385***	0.269***
	(0.062)	(0.048)	(0.041)	(0.073)	(0.06)	(0.05)
100 to 200 km	0.072	0.083*	0.017	0.342***	0.290***	0.114*
	(0.058)	(0.046)	(0.038)	(0.079)	(0.067)	(0.059)
200 to 300 km	0.138*	0.153***	0.065	0.210**	0.184***	0.058
	(0.072)	(0.059)	(0.049)	(0.082)	(0.069)	(0.064)
R-squared	0.289	0.417	0.323	0.259	0.38	0.306
Kleinberg-Paap F Stat	30.4	30.4	30.4	14.4	14.4	14.4
Median, interaction variable	0.444	0.444	0.444	0.725	0.725	0.725
Share of observations $\leq$ median						
0 to 100 km	0.483	0.483	0.483	0.588	0.588	0.588
100 to 200 km	0.49	0.49	0.49	0.48	0.48	0.48
200 to 300 km	0.526	0.526	0.526	0.449	0.449	0.449

Note: Trade at ports within 300 km is measured as of 1955. Highway miles measured c. 1960. Rail measured c. 1957. All regressions have 3,023 observations.

A Data Appendix

A.1 Data Sources

We use data from a variety of sources. This appendix provides source information.

1. County Business Patterns
   These data include total employment, total number of establishments (with some variation in this definition over time), and total payroll.

   - 1956: Courtesy of Gilles Duranton and Matthew Turner. See Duranton et al. (2014) for source details. We collected a small number of additional counties that were missing from the Duranton and Turner data.
     - In these data, payroll is defined as the "amount of taxable wages paid for covered employment [covered by OASI, or almost all "nonfarm industrial and commercial wage and salary employment" (page VII)[18]] during the quarter. Under the law in effect in 1956, taxable wages for covered employment were all payments up to th first $4,200 paid to any one employee by any one employer during the year, including the cash value of payments in kind. In general, all payments for covered employment in the first quarter were taxable unless the employee was paid at the rate of more than $16,800 per year. For the first quarter of 1956, it is estimated that 97.0 percent of total non-agricultural wages and salares in covered employment was taxable. The taxable proportion of total wages becomes smaller in the later quarter of the year. Data are presented for the first quarter because wages for this quarter are least affected by the provisions of the law limiting taxable wages to $4,200 per year." (page VI, Section III, Definitions in 1956 County Business Patterns report.)
   - 1967 to 1985: U.S. National Archives, identifier 313576.
   - 1986 to 2011: U.S. Census Bureau. Downloaded from `https://www.census.gov/econ/cbp/download/`
     - For comparability, we also use total first quarter payroll from these data.

2. Decennial Census: Population and demographics data by county

   - 1910: ICPSR 02896, Historical, Demographic, Economic and Social Data: The United States, 1790-2002, Dataset 38: 1950 Census I (County and State)
   - 1920: ICPSR 02896, Historical, Demographic, Economic and Social Data: The United States, 1790-2002, Dataset 38: 1950 Census I (County and State)
   - 1930: ICPSR 02896, Historical, Demographic, Economic and Social Data: The United States, 1790-2002, Dataset 38: 1950 Census I (County and State)

[18] Data also exclude railroad employment.

- 1940: ICPSR 02896, Historical, Demographic, Economic and Social Data: The United States, 1790-2002, Dataset 38: 1950 Census I (County and State)
- 1950
  - ICPSR 02896, Historical, Demographic, Economic and Social Data: The United States, 1790-2002, Dataset 38: 1950 Census I (County and State)
  - Census of Population, 1950 Volume II, Part I, Table 32.
- 1960: ICPSR 02896, Historical, Demographic, Economic and Social Data: The United States, 1790-2002, Dataset 38: 1960 Census I (County and State)
- 1970: ICPSR 8107, Census of Population and Housing, 1970: Summary Statistic File 4C – Population [Fourth Count]
- 1980: ICPSR 8071, Census of Population and Housing, 1980: Summary Tape File 3A
- 1990: ICPSR 9782, Census of Population and Housing, 1990: Summary Tape File 3A
- 2000: ICPSR 13342, Census of Population and Housing, 2000: Summary File 3
- 2010: U.S. Census Bureau, 2010 Decennial Census Summary File 1, Downloaded from `http://www2.census.gov/census_2010/04-Summary_File_1/`

3. Port Universe and Depth
   - We use these documents to establish the population of ports in any given year.
   - 1953: National Geospatial Intelligence Agency (1953)
   - 2015: National Geospatial Intelligence Agency (2015)
4. Port Containerization Adoption Year
   - 1956–2010: *Containerisation International Yearbook* for 1968 and 1970–2010
5. Port Volume: Total imports and exports by port
   - 1948: United States Foreign Trade, January-December 1949: Water-borne Trade by United States Port, 1949, Washington, D.C.: U.S. Department of Commerce, Bureau of the Census. FT 972.
   - 1955: United States Waterborne Foreign Trade, 1955, Washington, D.C. : U.S. Dept. of Commerce, Bureau of the Census. FT 985.
   - 2008: Containerisation International yearbook 2010, pp. 8–11.
6. Highways

- 2014: 2014 National Transportation Atlas, Office of the Assistant Secretary for Research and Technology, Bureau of Transportation Statistics, United States Department of Transportation. http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_atlas_database/2014/index.html.
- c. 1960: Office of Planning, Bureau of Public Roads, US Department of Commerce, "The National System of Interstate and Defense Highways." Library of Congress Call number G3701.P21 1960.U5. Map reports improvement status as of December 31, 1960.

7. Railways
   - 2014: 2014 National Transportation Atlas, Office of the Assistant Secretary for Research and Technology, Bureau of Transportation Statistics, United States Department of Transportation. http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_atlas_database/2014/index.html.
   - c. 1957: Army Map Service, Corps of Engineers, US Army, "Railroad Map of the United States," prepared 1935, revised April 1947 by AMS. 8204 Edition 5-AMS. Library of Congress call number G3701.P3 1957.U48.
8. Waterways
   - 2014: 2014 National Transportation Atlas, Office of the Assistant Secretary for Research and Technology, Bureau of Transportation Statistics, United States Department of Transportation. http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_atlas_database/2014/index.html.
9. World Urbanization Prospects: The 2014 Revision
   These data include population counts for all urban agglomerations whose populations exceed 300,000 at any time between 1950 and 2010.
   - Produced by the United Nations, Department of Economic and Social Affairs, Population Division.
   - Downloaded from http://esa.un.org/unpd/wup/CD-ROM/WUP2014_XLS_CD_FILES/WUP2014-F22-Cities_Over_300K_Annual.xls

A.2 Data Choices

1. U.S. County Sample

Our unit of analysis is a consistent-border county from 1950 to 2010. We generate these counties by aggregating 1950 counties. Please see the final Appendix Table for the specific details of aggregation.[19]

The 1956 County Business Patterns allowed for reporting of only 100 jurisdictions per state, leading to the reporting of aggregate values for agglomerations of counties in states with many counties. See Duranton et al. (2014) for the initial collection of these data, and additional details. To resolve the problem of making these 1956 units consistent with the 1950 census units, we disaggregate the 1956 CBP data in the agglomerated reporting into individual counties, attributing economic activity by population weights.

Alaska and Hawaii were not states in 1950. We omit Alaska from our sample, because in 1950 it has only judicial districts, which do not correspond to modern counties. We keep Hawaii, where the 1950 borders are relatively equivalent to modern counties. We also keep Washington, DC, in all years.

We also make a few additional deletions

- Two counties that only appear in the data (1910-1930) before our major period of analysis: Campbell, GA (13/041) and Milton, GA (13/203).
- Two problematic counties. Menominee, WI (55/078) created in 1959 out of an Indian reservation; it has very few people. Broomfield, CO (08/014), created in 2001 from parts of four other counties.
- Two counties where land area changes are greater than 40 percent. These are Denver County, CO (08/031) and Teton County, WY (56/039).

2. Ports

   We determine the universe of ports from the 1953 and 2015 *World Port Index*.

3. County Business Patterns data

   - For some county/industry groupings, there is a disclosure risk in reporting either the total number of employees or the total payroll. In such cases, we convert the disclosure code ("D" in the years before 1974) to 0.
   - "Payroll" is always first quarter payroll.

4. Income distribution calculations

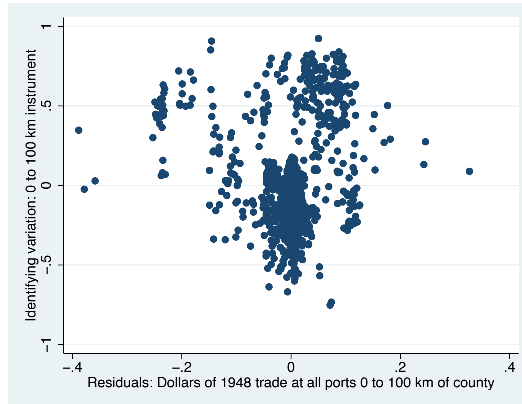
   - We use binned income data. In 1950, the number in each bin is the total number of families and unrelated individuals. In 2010, the number in each bin is to fix.

[19]These groupings relied heavily on the very helpful work of the Applied Population Laboratory group at the University of Wisconsin. See their documentation at `http://www.netmigration.wisc.edu/datadictionary.pdf`.

- To calculate percentiles, we assume that income is uniformly distributed within bins, with the exception of the top bin, which has no top code.
- For the top bin, we assume that income is distributed following a Pareto distribution, with a parameter $alpha$. We assume that $\alpha = \max(\hat{\alpha}, 1)$. Let N_B be the number of people in the top income bin, and N_{B-1} be the number of people in the second highest bin. Similarly, L_B be the lower bound of the top income bin and L_{B-1} be the lower bound of the second highest income bin. Then
$$\hat{\alpha} = \frac{\log(N_B + N_{B-1}) - \log(N_B)}{L_B - L_{B-1}}$$
.

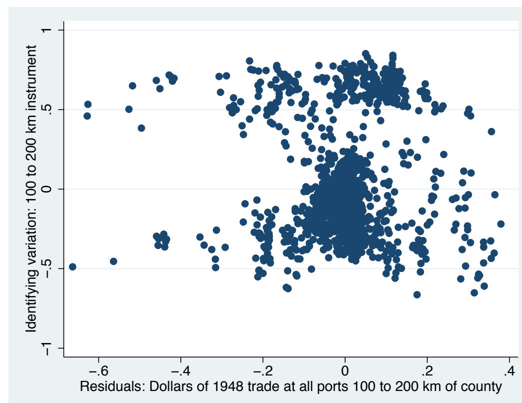
Appendix Figure 1: Instrument Variation vs. Pre-Treatment Covariates: All Instruments

(a)

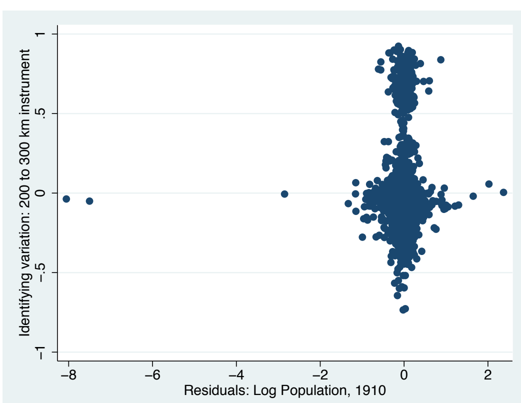


(b)

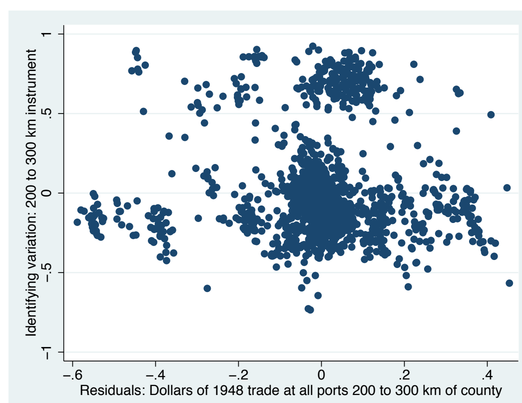
(c)



(d)



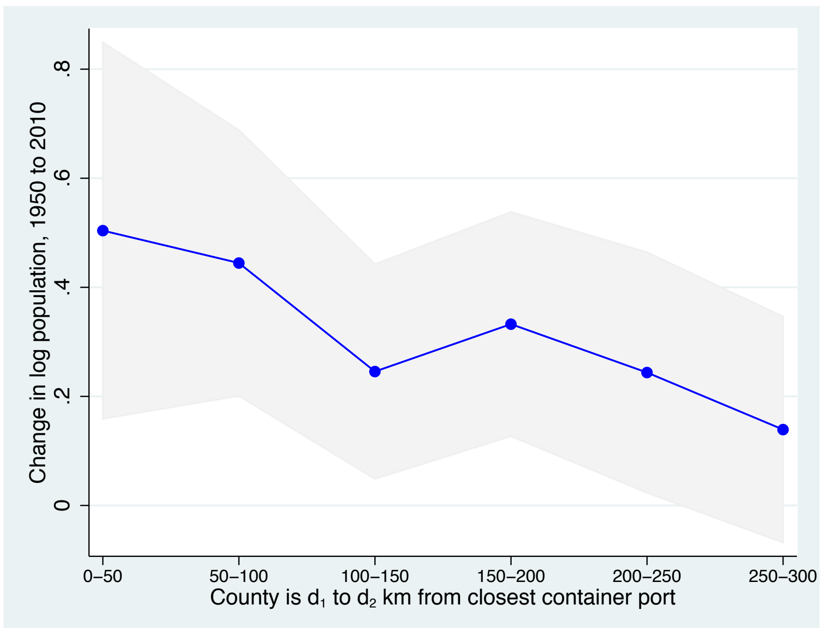
(e)



(f)

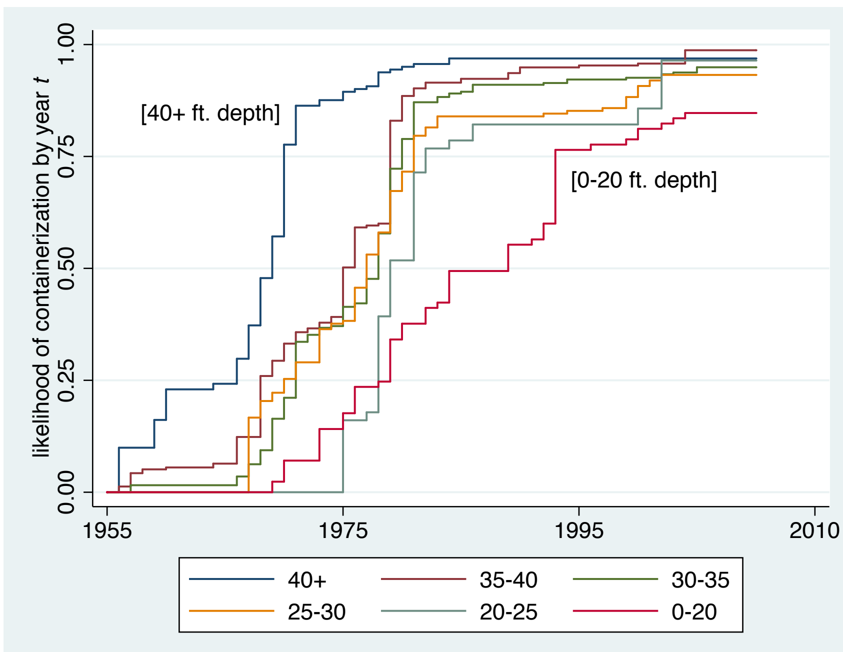
Notes: "Identifying variation" is the residual from a regression of the instrument (e.g. county is within d_1 to d_2 kilometers of a "very deep" port in 1953) on the full set of covariates. Appendix Figures 1a, 1c, and 1e plot the identifying variation versus 1910 log population, conditional on covariates (we use a regression to generate the relevant residual). Appendix Figures 1b, 1d, and 1f plot the identifying variation versus total dollars of 1948 international trade at ports between d_1 to d_2, conditional on covariates.

Appendix Figure 2: IV Estimates Indistinguishable From Zero at 300 km



Notes: This picture reports coefficients from the specification in column 8 of Table 2, but with six, rather than three, bins. Specifically, we paramaterize $\Delta C_{i,t}$ as six indicator variables, one for each distance bin of {0 to 50, 50 to 100, 100 to 150, 150 to 200, 200 to 250, 250 to 300} km. Each dot is the estimated coefficient for each indicator. Gray bands portray the 95% confidence interval.

Appendix Figure 3: World: Likelihood of Having a Containerized Port by 1953 Port Depth



Notes: In this picture, a city is containerized if it is within 300 km of a container port and $t >$ year of first containerization of any port within 300 km. Port depth is the depth of the deepest wharf within 300 km in 1953. Deeper ports are more likely to ever containerize, and more likely to containerize early.

Appendix Table 1: Complete First Stage Specification

	1 if Closest Container Port is d_1 to d_2 km of county		
	0 to 100	100 to 200	200 to 300
	(1)	(2)	(3)
County is d_1 to d_2 of a very deep port			
0 to 100 km	0.530***	0.062	-0.017
	(0.038)	(0.039)	(0.029)
100 to 200 km	0.004	0.602***	-0.01
	(0.02)	(0.03)	(0.025)
200 to 300 km	-0.024	-0.019	0.637***
	(0.016)	(0.025)	(0.031)
R-squared	0.577	0.461	0.415
Joint F test, instruments	72.1	157.6	159.8
Mean, dependent variable	0.122	0.173	0.146

Notes: All estimations use 3,023 observations. Stars denote significance levels: * 0.10, ** 0.05, and *** 0.01. The F test values in this table are from a test of joint significance of the three reported coefficients. Table 2 reports the Kleinberg-Paap F statistic, as suggested by Sanderson and Windmeijer (2016)

Appendix Table 2: Containerization Yields Shifts Toward More Foreign, Educated and Younger Population

	IV, Dependent Variable is					
	Foreign born (share)	Age 65 and older (share)	People $\geq$ 25, $<$ high school degree (share)	Log of 90/10 income ratio	Log of 90/50 income ratio	Log of 50/10 income ratio
	(1)	(2)	(3)	(4)	(5)	(6)
Closest container port is						
0 to 100 km	0.066***	-0.036***	-0.023*	-0.018	-0.138***	0.120**
	(0.010)	(0.007)	(0.012)	(0.049)	(0.029)	(0.041)
100 to 200 km	0.054***	-0.012*	-0.005	-0.006	-0.037	0.031
	(0.006)	(0.005)	(0.009)	(0.038)	(0.023)	(0.030)
200 to 300 km	0.030***	-0.011*	-0.011	-0.069+	-0.065**	-0.004
	(0.006)	(0.005)	(0.008)	(0.039)	(0.024)	(0.029)
R-squared	0.288	0.155	0.535	0.198	0.352	0.184
Kleinberg-Paap F Stat	70.238	70.238	70.238	70.094	70.094	70.094
Mean, Dependent Variable	0.015	0.072	-0.584	-0.371	0.028	-0.399
Observations	3023	3023	3023	3022	3022	3022

Notes: These specifications are all IV regressions including the most complete covariate list from Table 2. Column 1, 2, and 3 use census demographics data. Column 4, 5, and 6 use census income data. Stars denote significance levels: * 0.10, ** 0.05, and *** 0.01.

Appendix Table 3: World City Characteristics by Distance to Containerized Port

	Distance to Containerized Port							
	0 to 50	50 to 100	100 to 150	150 to 200	200 to 250	250 to 300	Ever Cont.	Never Cont.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Population								
1950	12.674	12.493	12.399	12.365	12.331	12.321	12.322	11.985
	[1.147]	[1.087]	[1.105]	[1.081]	[1.003]	[1.028]	[1.058]	[0.811]
2010	14.112	13.799	13.722	13.666	13.653	13.666	13.810	13.603
	[1.067]	[1.026]	[0.999]	[0.962]	[0.903]	[0.917]	[0.978]	[0.804]
Continent								
Africa	0.101	0.055	0.056	0.100	0.055	0.067	0.097	0.048
Asia	0.348	0.346	0.372	0.363	0.376	0.412	0.378	0.585
Australia	0.033	0.011	0.017	0.012	0.000	0.011	0.014	0.000
Europe	0.272	0.313	0.325	0.327	0.345	0.326	0.241	0.186
North America	0.156	0.203	0.175	0.143	0.188	0.139	0.184	0.115
South America	0.091	0.071	0.056	0.056	0.035	0.045	0.087	0.067
Years Since	32.438	31.852	30.774	30.378	31.176	30.899	35.171	.
First Cont.	[11.158]	[11.790]	[12.406]	[11.845]	[12.125]	[11.513]	[9.667]	.
Observations	276	182	234	251	255	267	632	419

The unit of observation in this table is the city.
We report means and standard deviations (brackets) for each variables.
Source: See data appendix.

Appendix Table 4: First Stage: Containerization More Likely When Ports are Deep (World Sample)

	Ever Cont.			Years of Cont.		
	(1)	(2)	(3)	(4)	(5)	(6)
	0 to 100	100 to 200	200 to 300	0 to 100	100 to 200	200 to 300
30+ ft. Depth (0-100km)	0.552***	-0.095**	-0.045	21.236***	-3.558***	-2.925**
	(0.036)	(0.038)	(0.040)	(1.413)	(1.328)	(1.397)
30+ ft. Depth (100-200km)	-0.086**	0.499***	-0.066	-2.895**	18.119***	-2.820**
	(0.037)	(0.036)	(0.040)	(1.338)	(1.425)	(1.366)
30+ ft. Depth (200-300km)	-0.106***	-0.096**	0.484***	-2.817*	-2.457*	19.454***
	(0.041)	(0.042)	(0.040)	(1.456)	(1.463)	(1.498)
Mean, Dependent Variable	0.35	0.35	0.37	11.71	11.08	12.02
R-squared	0.65	0.63	0.63	0.68	0.66	0.67
F Stat Excluded Instrument(s)	99.72	73.97	62.39	97.77	64.31	84.79

Notes: Sample of world cities over 50K in 1950. Dependent variable (1)-(3): Adoption of containerization between d1 and d2 km. Dependent variable (4)-(6): Number of years since the adoption of containerization between d1 and d2 km. All specifications control for country fixed effects, log population in 1950, a dummy variable for having a port within 300km in 1953, and the number of ports within 300km in 1953. All regressions have 1051 observations. Port depth is the depth of the deepest port within 300km. Robust standard errors in parentheses. Stars denote significance levels: * 0.10 ** 0.05 *** 0.01. Sources: See data appendix.

Appendix Table 5: County Groupings for Consistent Counties

State	State FIPS	Grouped County FIPS	Initial Counties: County Name	Initial Counties: County FIPS	Notes
Arizona	04	027	La Paz County	012	Used to be part of Yuma County (04/027)
Florida	12	086	Miami Dade	025	Name change, from Dade County to Miami-Dade, yielded a numbering change.
Hawaii	15	010	Kalawao County	005	
Hawaii	15	010	Maui County	009	
Montana	30	067	Yellowstone County	113	Yellowstone County merged is to Park County (30/067)
Nevada	32	510	Ormsby County	025	Becomes Carson City (32/510)
New Mexico	35	061	Cibola County	006	Used to be part of Valencia County (35/061)
South Dakota	46	041	Armstrong County	001	Is merged into Dewey County (46/041)
South Dakota	46	071	Washabaugh County	131	Is merged into Jackson County (46/071)
Virginia	51	900	Albermarle County	003	
Virginia	51	901	Alleghany County	005	
Virginia	51	906	Arlington County	013	
Virginia	51	902	Augusta County	015	
Virginia	51	903	Bedford County	019	
Virginia	51	903	Campbell County	031	
Virginia	51	904	Carroll County	035	
Virginia	51	905	Chesterfield County	041	
Virginia	51	915	Dinwiddie County	053	
Virginia	51	924	Elizabeth City	055	

Virginia	51	906	Fairfax County	059	
Virginia	51	907	Frederick Couty	069	
Virginia	51	904	Grayson County	077	
Virginia	51	908	Greensville County	081	
Virginia	51	909	Halifax County	083	
Virginia	51	905	Henrico County	087	
Virginia	51	910	Henry County	089	
Virginia	51	911	James City County	095	
Virginia	51	912	Montgomery County	121	
Virginia	51	800	Nanasemond City	123	Is later folded into Suffolk County (51/800)
Virginia	51	913	Norfolk County	129	
Virginia	51	914	Pittsylvania County	143	
Virginia	51	915	Prince George County	149	
Virginia	51	913	Princess Anne	151	
Virginia	51	916	Prince William County	153	
Virginia	51	917	Roanoake County	161	
Virginia	51	918	Rockbridge County	163	
Virginia	51	919	Rockingham County	165	
Virginia	51	920	Southhampton County	175	
Virginia	51	921	Spotsylvania County	177	
Virginia	51	924	Warwick County	189	
Virginia	51	922	Washington County	191	
Virginia	51	923	Wise County	195	
Virginia	51	924	York County	199	
Virginia	51	906	Alexandria City	510	
Virginia	51	903	Bedford City	515	
Virginia	51	922	Bristol City	520	
Virginia	51	918	Buena Vista City	530	
Virginia	51	900	Charlottesville City	540	
Virginia	51	913	Chesapeake City	550	
Virginia	51	901	Clifton Forge City	560	
Virginia	51	905	Colonial Heights City	570	
Virginia	51	901	Covington City	580	

Virginia	51	914	Danville City	590	
Virginia	51	908	Emporia City	595	
Virginia	51	906	Fairfax City	600	
Virginia	51	906	Falls Church City	610	
Virginia	51	920	Franklin City	620	
Virginia	51	921	Fredricksburg City	630	
Virginia	51	904	Galax City	640	
Virginia	51	924	Hampton City	650	
Virginia	51	919	Harrisonburg City	660	
Virginia	51	915	Hopewell City	670	
Virginia	51	918	Lexington City	678	
Virginia	51	903	Lynchburg City	680	
Virginia	51	916	Manassas City	683	
Virginia	51	916	Manassas Park City	685	
Virginia	51	910	Martinsville City	690	
Virginia	51	800	Nanasemond County	695	Appears for a few years in County Business Patterns data as a county.
Virginia	51	924	Newport News City	700	
Virginia	51	913	Norfolk City	710	
Virginia	51	913	Portsmouth City	710	
Virginia	51	923	Norton City	720	
Virginia	51	915	Petersburg City	730	
Virginia	51	924	Poquoson City	735	
Virginia	51	912	Radford City	750	
Virginia	51	905	Richmond City	760	
Virginia	51	917	Roanoake City	770	
Virginia	51	917	Salem City	775	
Virginia	51	909	South Boston City	780	
Virginia	51	913	South Norfolk City	785	
Virginia	51	902	Staunton City	790	
Virginia	51	913	Virginia Beach City	810	
Virginia	51	902	Waynesboro City	820	
Virginia	51	911	Williamsburg City	830	

Virginia	51	907	Winchester City	840	
Wyoming	56	039	Yellowstone Park County	047	Is merged into Teton County (56/039)