Investment in Infrastructure and Trade: The Case of Ports

Giulia Brancaccio[*], Myrto Kalouptsidi[†], Theodore Papageorgiou[‡§]

New York University, *Harvard University,* *Boston College*

May 2024

Returns to Investing in Port Infrastructure

- 80% of trade carried by ship (but not that much bulk!)
- Three policy relevant messages
  - investing in port infrastructure can lead to substantial trade and welfare gains if targeted properly
  - there are sizable spillovers across ports (one port can decongest a wider set of ports)
  - (iii) macroeconomic volatility can drastically change the returns to investment and their geography.

	Mean	SD	25th percentile	75th percentile
Annual number of port calls	136	130	54	180
Annual throughput (tons)	7,295,140	7,355,597	2,602,124	9,970,174
Time at port (hours): average	117	54	74	153
Time at port (hours): st. dev. over time	88	41	58	118
Wait time (hours): average	39	39	2.9	58
Wait time (hours): st. dev. over time	58	50	16	101
Fraction of port calls with positive wait time	41%	30%	13%	70%
Port dues ($ per ton)	2.1	1.3	1.2	2.9
Port dues (in $) per port call	117,345	96,206	44,038	168,399
Length of berths (miles)	0.96	0.8	0.42	1.4
Storage space (acres)	176	176	55	234

Table 1: Summary Statistics. Port calls (throughput in tons) is the total annual number of ships (total tonnage) that were handled at port. The average time at port refers to time at port averaged over time, while standard deviation of time at port is computed over time per port. The fraction of port calls with positive wait time is computed for each port and year.

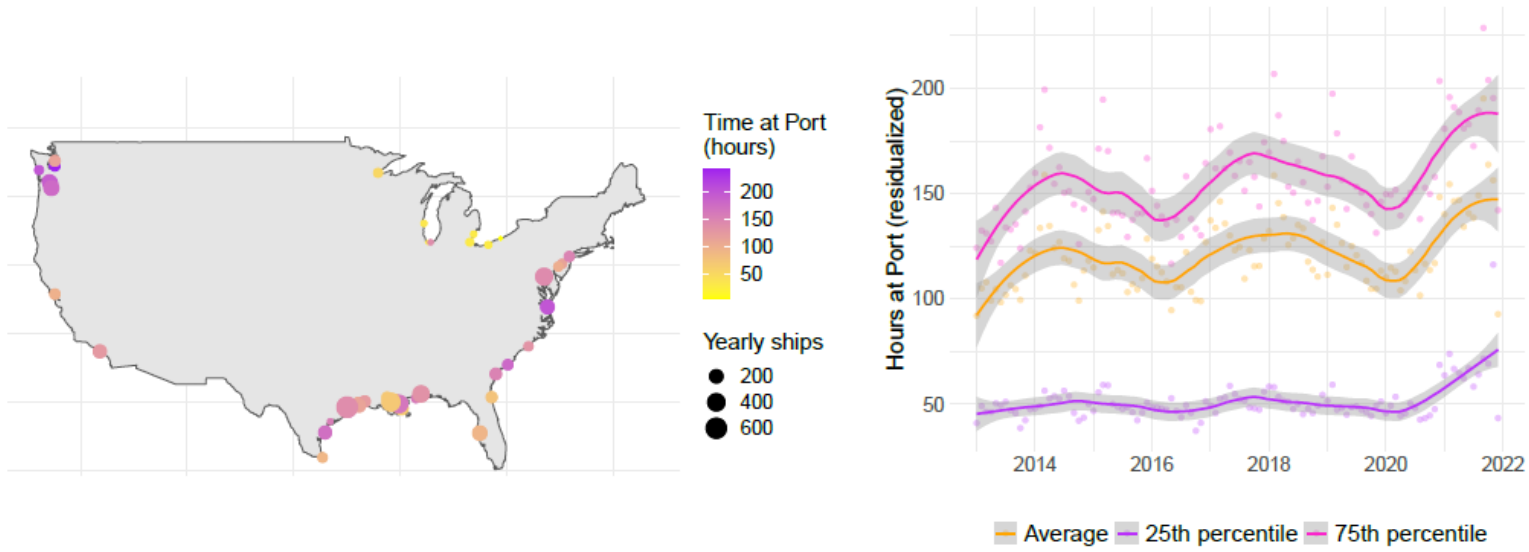

Figure 1: Port throughput and time at port over space and time. The left panel plots the average time at port in US ports. It also indicates the size of the port in terms of annual throughput, designated by the size of its dot. The right panel plots the average time at port for each month, as well as its 25th and 75th percentile. We control for ship size and commodity on board.

handle at most K ships at a time. Then the ship's expected total time at port is equal to,

$$\underbrace{T}_{\text{service time}} + \mathbb{I}\{Q \geq K\} \underbrace{(Q - K + 1)\frac{T}{K}}_{\text{wait time (queueing)}} \tag{1}$$

In words, if the ship arrives and there are fewer than K ships at port (i.e. $Q < K$), then the port is operating under capacity, and the ship is able to be serviced immediately, so that total time at port simply equals the service time, T. If instead $Q \geq K$, then the ship needs to wait for a berth to open up and the total time is augmented by the wait time.

To understand the expression for wait time in equation (1), note that if the port could handle only one ship at a time ($K = 1$), then the expected wait time would just equal QT, as the ship would need to wait for Q ships to complete service, each taking on average T periods. In the case where $K > 1$, the arriving ship does not have to wait for all Q ships ahead to complete service, since the port can handle more than one ship at a time. For example, if $Q = K$, then only one ship needs to complete service for a berth to open up, regardless of how large Q is. In general, $Q - K + 1$ ships need to conclude service before the ship can get to berth.[13] Since service times are exponentially distributed, a berth becomes available in T/K periods in expectation when K ships are being serviced. Therefore the expected wait time is $(Q - K + 1)(T/K)$.[14]

Expected service time at port is given by

$$T(L/s, c/s, \omega) \tag{2}$$

and depends on the port's workers, L, cranes c, productivity ω, as well as the number $s \leq K$ of ships currently serviced. This specification is driven by the empirical finding that a ship's service time depends on how many other ships are being serviced (see Table 5 in Appendix A), suggesting that ports to some extent stretch their inputs across ships handled.[15]

Finally, we discuss how the port queue evolves endogenously as ships arrive, are serviced, and leave. To this end, denote by λ the ship arrival rate and by $\mu(s) = 1/T(s)$, the rate at which a ship completes service, where we suppress the dependence of $T(s)$ on L, c, ω.[17] Then, the evolution of the queue follows a Markov structure. Indeed the probability $p_n(\tau)$ that in instant τ there are n total ships at port, either being serviced or waiting, satisfies,

$$\frac{dp_n(\tau)}{d\tau} = \begin{cases} -\lambda p_0(\tau) + \mu(0)\, p_1(\tau), & \text{if } n = 0 \\ -(\lambda + n\mu(n))\, p_n(\tau) + \lambda p_{n-1}(\tau) + (n+1)\,\mu(n) p_{n+1}(\tau), & \text{if } 0 < n < K \\ -(\lambda + K\mu(K))\, p_n(\tau) + \lambda p_{n-1}(\tau) + K\mu(K) p_{n+1}(\tau), & \text{if } n \geq K \end{cases} \tag{3}$$

In words, consider the case where $n < K$ (a similar argument hold for $n \geq K$), then the change in the probability that there are n ships at port increases by (i) the probability that there are $n-1$ ships at port at τ, $p_{n-1}(\tau)$, and a new ship arrives, which happens with probability λ; and (ii) the probability that there are $n+1$ ships at τ, $p_{n+1}(\tau)$, and one of them is serviced, which happens with probability $(n+1)\,\mu(n)$. In addition the probability $p_n(\tau)$ decreases by the probability that there are n ships at port, $p_n(\tau)$, and a new ship arrives or one is serviced. Therefore, simulating the queue evolution boils down to solving this system of differential equations. We describe these simulations in detail in Appendix C.[18]

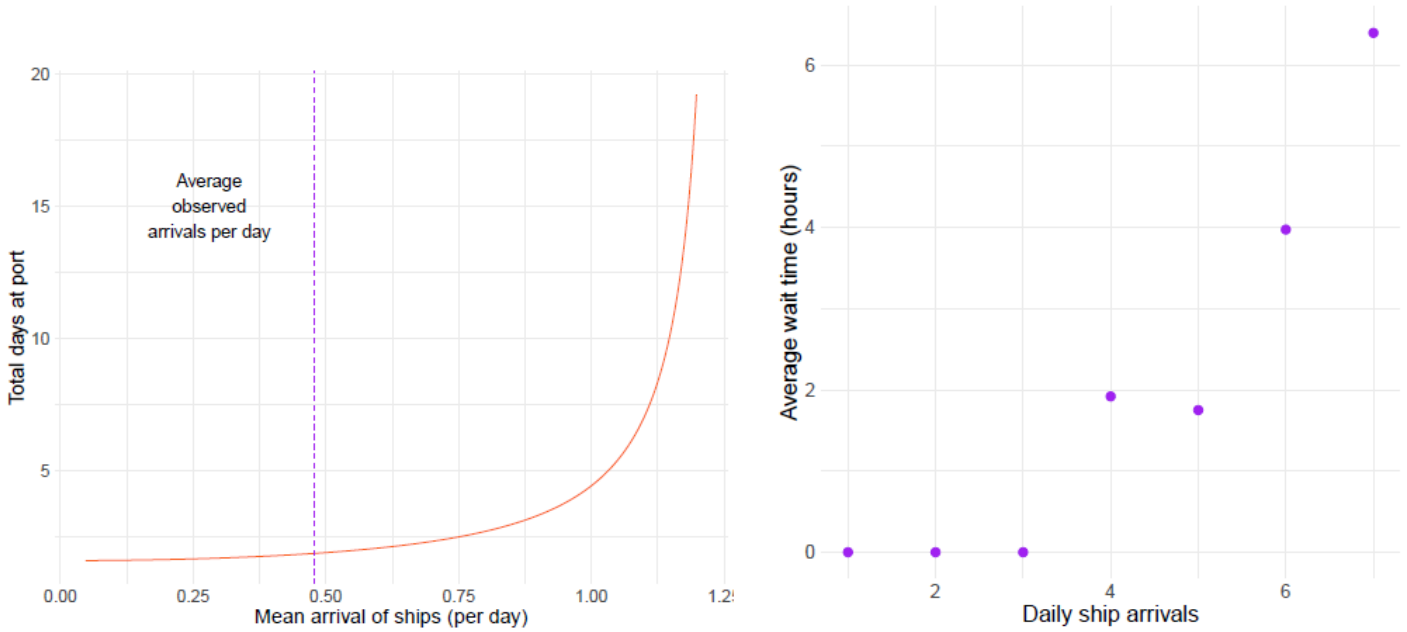

Figure 2: Convex cost of congestion. The left panel of this graph plots the relationship between time at port (vertical axis) and the mean arrival rate (horizontal axis) and is produced from model simulations of increasing the mean arrival rate of ships. In particular, for each level of the mean arrival rate on the horizontal axis, we simulate the port queue using the system of differential equations (3). We use the average service time, and the average capacity K. The right panel plots the empirical counterpart of this convex cost of congestion in the port of Duluth (MN). In particular, it plots the average time ships spend waiting (vertical axis) against a moving average of daily ship arrivals (horizontal axis).

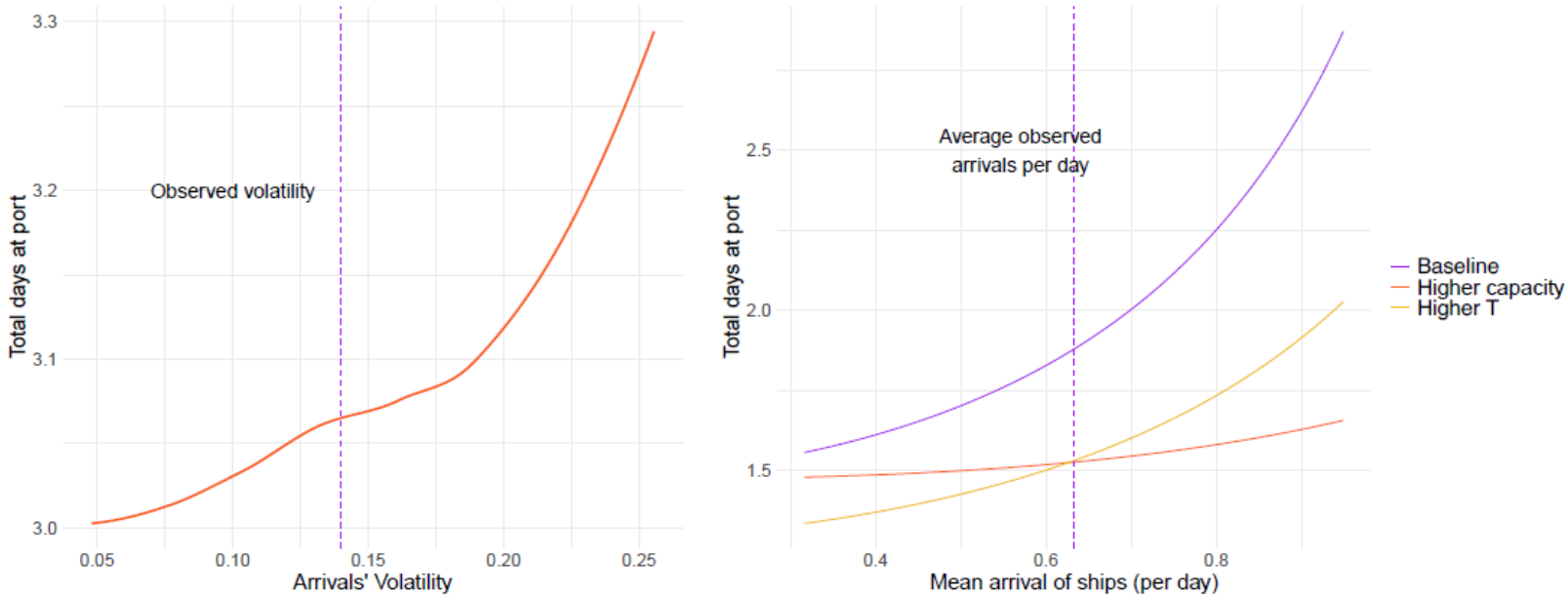

Figure 3: Congestion dynamics: the role of volatility and different port inputs. The left panel plots the relationship between time at port (vertical axis) and the variance of the arrival rate (horizontal axis). It is produced from model simulations of increasing mean preserving spreads of the arrival rate of ships. The right panel plots the relationship between time at port (vertical axis) and mean arrival rate (horizontal axis) when service time T falls or capacity K increases.

3.2 Demand for Ports

We next turn to demand for port services, which is necessary to estimate the welfare gains from port improvements.

There are D locations in the US, including both coastal and inland regions, and F locations abroad. For every pair (d, f) of US (domestic) and foreign locations, every month there are M_{df} importer-exporter pairs that have the option of trading a ship-sized shipment between the two locations, either originating in d, or destined there. In the case of exports, d is a US producing location and f is a foreign destination, while in the case of imports, d is a US importing destination and f is a foreign producing location. Let t denote the month-year period. Each shipment i needs to decide whether to trade and, if so, which port, $j \in \{1, ..., J\}$ to use. The payoff to shipment i, traveling between d and f, from choosing port j in period t is given by,

$$u_{ijt} = \beta_f \text{dist}(f, j) + \beta_d \text{dist}(j, d) - \beta_T TT_{jt} - \beta_p p_{jt} + \gamma_t + \gamma_f + \gamma_{l(j)} + \xi_{jft} + \epsilon_{ijt} \tag{4}$$

where $\text{dist}(f, j)$ denotes the distance between the foreign location f and port j and$\text{dist}(j, d)$ is analogously defined, TT_{jt} is the total time at port j in period t, and p_{jt} is the price that port j charges for its services in period t. The β parameters capture how much the pair values each characteristic. Moreover, γ_t is a month fixed effect capturing macroeconomic fluctuations, while γ_f and $\gamma_{l(j)}$ are respectively fixed effects for the foreign location and the port's region (namely, either the Gulf, the East Coast, the West Coast,

Trade flows The payoff specification (4) implies that the probability that a pair trading between foreign location f and domestic location d chooses port j is,

$$\sigma_{jfdt} \equiv \Pr(j|f,d,t) = \frac{\exp(\delta_{jft} + \beta_d \text{dist}(j,d))}{1 + \sum_l \exp(\delta_{lft} + \beta_d \text{dist}(l,d))} \tag{5}$$

where following Berry et al. (1995) we define,

$$\delta_{jft} \equiv \beta_f \text{dist}(f,j) - \beta_T TT_{jt} - \beta_p p_{jt} + \gamma_t + \gamma_f + \gamma_{l(j)} + \xi_{jft}$$

which captures the "mean utility" of port j for shipments from/to f in month t. The probability of the outside option, σ_{0fdt}, is similarly defined. This implies that every month the total flow through j is given by $\sum_f \sum_d \sigma_{jfdt} M_{df}$.

Looking ahead at the estimation, one complication of this framework is that our data on ship flows, do not contain information on a shipment's US location d, so we do not directly observe $\Pr(j|f,d,t)$. In other words, we might know that a ship began from China and unloaded at the port of Long Beach (CA), but we do not know the shipment's final destination within the US; similarly in the case of exports, we know that a shipment loaded in Houston (TX) and traveled to Europe, but we do not know where in the US it was originated.

To overcome this challenge, we treat the domestic location d as a latent variable that needs to be estimated (this is reminiscent of unobserved consumer types as in Berry et al., 2006). In particular, marginalizing (5) by the probability, $\Pr(d|f,t)$, of a shipment to/from any domestic US location d given the foreign location f, we obtain,

$$\sigma_{jft} \equiv \Pr(j|f,t) = \sum_{d} \Pr(j|f,d,t) \Pr(d|f,t) \tag{6}$$

Note that $\Pr(j|f,t)$ is observed in our ship flow data. In order to measure $\Pr(d|f,t)$, we leverage the FAF Regional Database, which contains information on freight movements among major US regions. In particular, the FAF reports yearly trade flows between US regions, such as states or metropolitan areas, as well as flows between these regions and broad foreign locations. For instance, it reports flows from California to Southeast Asia or Europe, which allows us to measure $\Pr(d|f,t)$. These moments will be useful in writing down our model's likelihood as we discuss in Section 4.2 below.

Finally, the specific timing of the $\sum_f \sum_d \sigma_{jfdt} M_{df}$ ships that pass through port j within month t is random. In particular, a ship arrives at port j at Poisson rate $\lambda_{jt} = \sum_f \sum_d \sigma_{jfdt} M_{df}$.

4.1 Port Technology Estimation and Results

Service Time We specify port service time as a general CES specification where

$$T_{jt} = \omega_{jt} \left(\alpha \left(\frac{L_{jt}}{s_{jt}} \right)^{\eta} + (1-\alpha) \left(\frac{c_{jt}}{s_{jt}} \right)^{\eta} \right)^{-\frac{1}{\eta}} \tag{8}$$

[21] In fact, it would have been equivalent to define the payoff of the outside option as $u_{i0t} = \gamma_t + \epsilon_{i0t}$. It is however easier in the estimation to subtract γ_t from the utility of all alternative options j, and set the mean payoff from the outside option equal to zero, so that $u_{i0t} = \epsilon_{i0t}$.

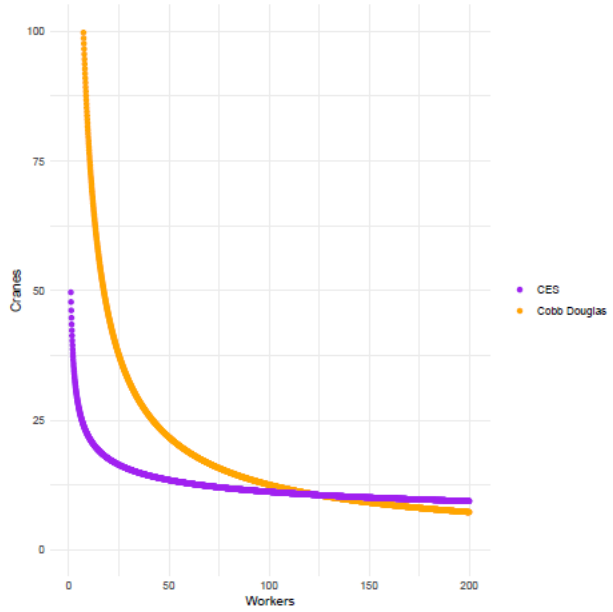

Figure 4: Port service time. The figure plots the isoquants associated with the estimated service time CES specification (8) (for the average service time); it also plots the corresponding isoquant if we were to estimate a Cobb-Douglas specification (i.e. $\eta \to 0$).

Following the usual concern when estimating production functions that inputs are not orthogonal to realized productivity, we introduce instruments, z_{jt}. First, we use local wages in related occupations as an instrument for labor. In particular, wage changes in occupations such as truck drivers, construction equipment operators, and taxi drivers, the most common substitute occupations for longshoremen (as indicated by occupational transition matrices), change their outside options and thus affect port employment. We construct this instrument using wage data from the Occupational Employment and Wage Statistics of the BLS and occupational switching probability data from the Current Population Survey (CPS) of the US Census Bureau. As additional shifters for a port's labor force, we use changes to the labor force and to the fraction of the population above 65 years old in the port county. Finally, we include lagged inputs as instruments. We leverage these instruments to estimate equation (8) using a GMM estimator based on the moments $\mathbb{E}\left(\nu_{jt} z_{jt}\right) = 0$.[22]

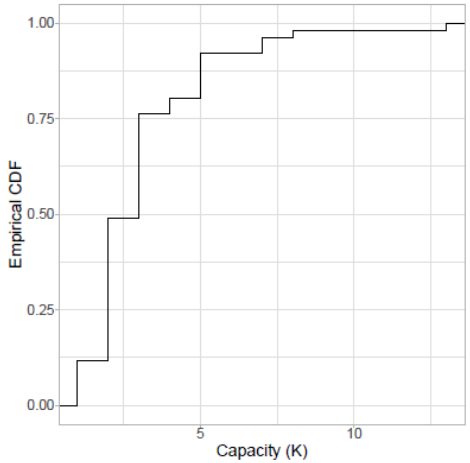

Figure 5: Port Capacity. This figure plots the empirical cumulative distribution of ports' effective capacity K_j, measured using AIS data, as the maximum number of ships that can be handled simultaneously, conditional on at least one ship waiting in anchorage.

We assume that capacity is given by a Leontief function: a ship requires both space to park (length of berth), as well as the land space to (un)load the cargo, and it is unlikely that the two are flexibly substitutable. Therefore we have,[24]

$$K_j = \min\left\{\frac{B_j}{\kappa_1}, \frac{A_j}{\kappa_2}\right\} \tag{9}$$

where B_j is port j's total length of its berths, A_j is its acreage of storage space, and κ_1, κ_2 are parameters.[25] In particular, κ_1 and κ_2 are technological constants determining, respectively, the mileage of berths and acreage of storage required to handle a ship.

Based on equation (6), we formulate a likelihood for the port call microdata as follows:[27]

$$\begin{aligned}\mathcal{L}^{\text{micro}}(\delta, \beta_d) &= \sum_{f,t}\sum_{j} n_{ft}^{j} \log\left(\Pr\left(j|f,t\right)\right) = \\ &= \sum_{f,t}\sum_{j} n_{ft}^{j} \log\left(\sum_{d} \Pr\left(j|f,d,t\right)\Pr\left(d|f,t\right)\right) \\ &= \sum_{f,t}\sum_{j} n_{ft}^{j} \log\left(\sum_{d} \frac{\exp\left(\delta_{jft} + \beta_d \text{dist}(j,d)\right)}{1+\sum_l \exp\left(\delta_{lft} + \beta_d \text{dist}(j,d)\right)} \Pr\left(d|f,t\right)\right),\end{aligned}$$

[24] We omit time variation, as these variables exhibit very little variation in our sample.

where n_{ft}^{j} is the number of ships that choose port j in period t with foreign location f. We denote by $M_f = \sum_d M_{df}$ the number of potential shippers in market f,t, which we measure using the maximum observed yearly flow in FAF from/to foreign location f.[28]

Note that total time at port may be correlated with other time-varying characteristics of the port captured by ξ_{jft}. For example, an increase in demand driven by higher ξ_{jft}, by construction, leads to higher time at port, TT_{jt}, as more ships now choose port j. For this reason, it is necessary to use an instrument for TT_{jt}. We use unexpected disruptions in port j's service operations at t, which increase the total time a ship spends at port j. These disruptions directly affect TT_{jt} and, through the increased time at port, also reduce demand for port j as fewer shippers now choose this port; at the same time however, these disruptions affect how many ships arrive only through their impact on TT_{jt}, rather than directly. In particular, we employ the residuals from a regression of service time on a number of controls, including labor, cranes and fixed effects; these residuals are meant to capture unexpected disruptions in service, such as crane breakdowns and strikes.

In order to estimate the demand model using the above instrument, we follow Grieco et al. (2022). Their procedure, aimed at environments with micro-data on individual choices, but where endogeneity must be addressed, amounts to augmenting the likelihood above by a set of moments based on the exogeneity of ξ_{jft} from instruments, of the form

$$\mathbb{E}\left(\xi_{jft} z_{jft}\right) = \mathbb{E}\left[\left(\delta_{jft} - \beta_f \text{dist}(f,j) + \beta_T TT_{jt} + \beta_p p_{jt} - \gamma_t - \gamma_f - \gamma_{l(j)}\right) z_{jft}\right] = 0 \tag{10}$$

In addition, in order to identify the parameter β_d, which measures the importance of the distance from the unobserved domestic location d, we further augment the above micro-data likelihood by what we call a "macro-likelihood". This matches (yearly) aggregate flows between every foreign location f and every domestic location d, through port j:

$$\mathcal{L}^{\text{macro}}\left(\delta, \beta_d\right) = \sum_{\text{year}} \sum_{f,d,j} \left(\sigma_{fjd,\text{year}}\right)^{n_{fjd,\text{year}}}$$

where $n_{fjd,\text{year}}$ is the annual flow from f to d through j. It is important to note that this term is based on the annual flows reported in FAF, so it is a substantially aggregated version of the port call microdata.[29]

Demand Estimates	
Average time at port	−1.44
	(0.49)
Port price per ton	−1.60
	(0.02)
Foreign distance (days of travel)	−0.71
	(0.15)
Domestic distance (log days of travel)	−1.2
	(0.05)
East Coast FE	-0.98
	(2.10)
Great Lakes FE	-15.89
	(4.57)
Gulf FE	-8.22
	(3.73)
Foreign location FE	Yes
Month FE	Yes
Instruments	Unexpected disruptions to port operations

Table 2: Demand Estimates. This table reports the estimated parameters of the payoff function (4). Standard errors are obtained from 500 bootstrap samples.

4.2.1 Results

Table 2 presents the estimated parameters. We begin by computing the "value of time" implied by our estimates, which is measured by the ratio β_T/β_p and captures the amount carriers are willing to pay in order to reduce time at port by one day. We find a value of time of \$0.9 per ton of deadweight, hence \$45,000 for the average ship.[32,33] This estimate for the value of time is substantial, especially compared to the direct cost of \$14,000 that the shipper needs to pay to hold the ship for an extra day. The difference between our estimate of value of time and this direct cost reflects the cost of port disruptions as they propagate through the supply chain.

In addition, we find that the elasticity with respect to time at port is large, as shown in Figure 6, suggesting that any disruptions affecting time at port are likely to have a large impact on demand. Moreover it is important to note that there is large heterogeneity in the elasticity across space which we return to below.

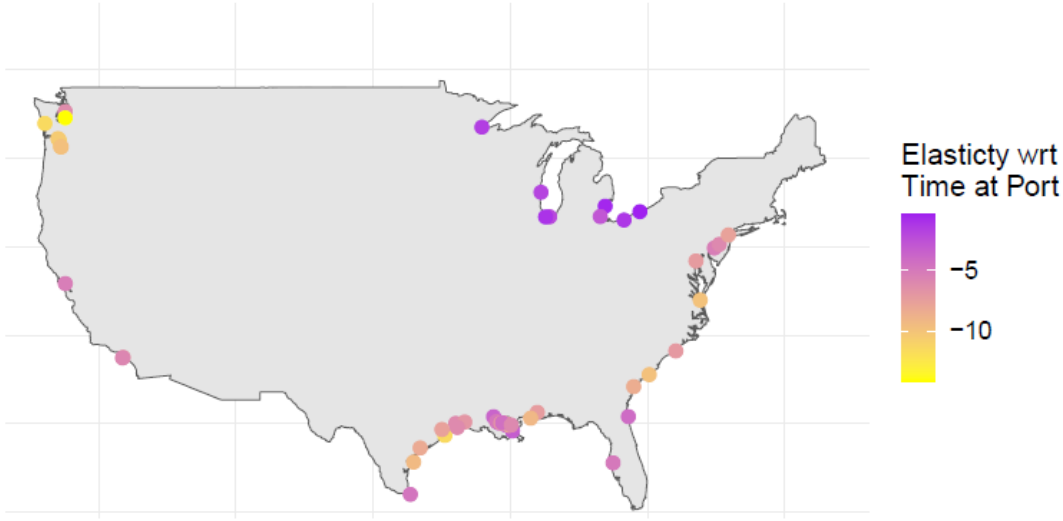

Figure 6: Demand Elasticity. This figure plots the average elasticity with respect to time at port for each US port.

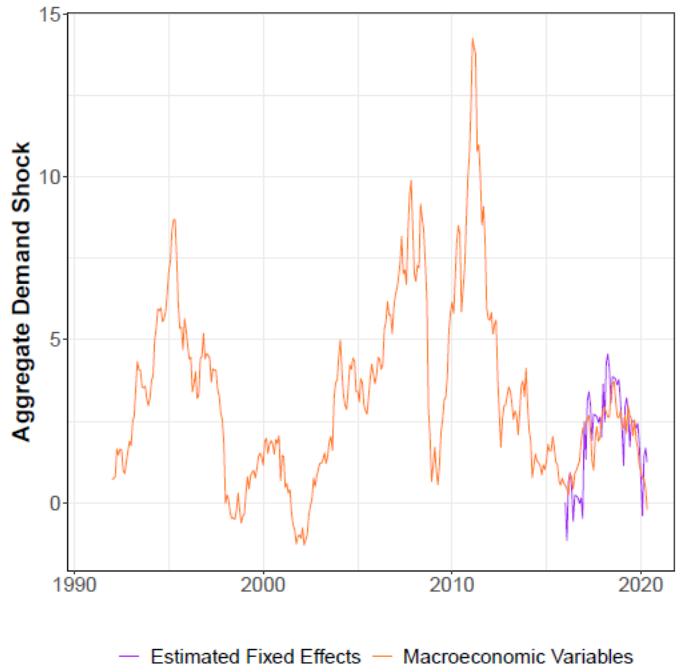

Figure 9: Macroeconomic volatility. This figure plots the estimated fixed effects, γ_t, as well as the predicted time fixed effects by macroeconomic variables, $\hat{\gamma}_t \equiv \hat{\beta} y_t$. The macroeconomic variables y_t include the Baltic Dry Index, commodity prices, and the industrial price index. See Table 9 in Appendix A for details.

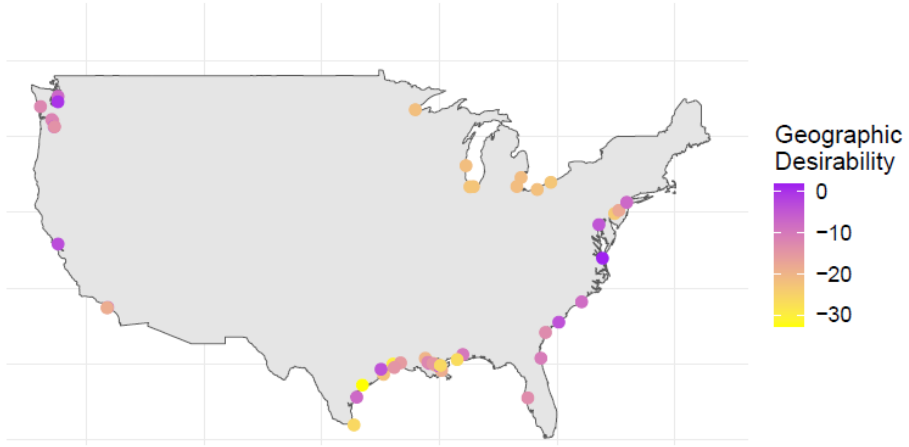

Figure 7: Geographic Attractiveness. This figure plots the variable $\delta^{\text{geo}}_{jfdt} = \beta_f \text{dist}(f,j) + \beta_d \text{dist}(j,d) + \gamma_{l(j)} + \xi_{jft}$, for each US port, averaged across domestic and foreign locations, as well as time t. Domestic and foreign locations are weighted by the market size of each (d, f) pair. This captures the geographic attractiveness of each port.

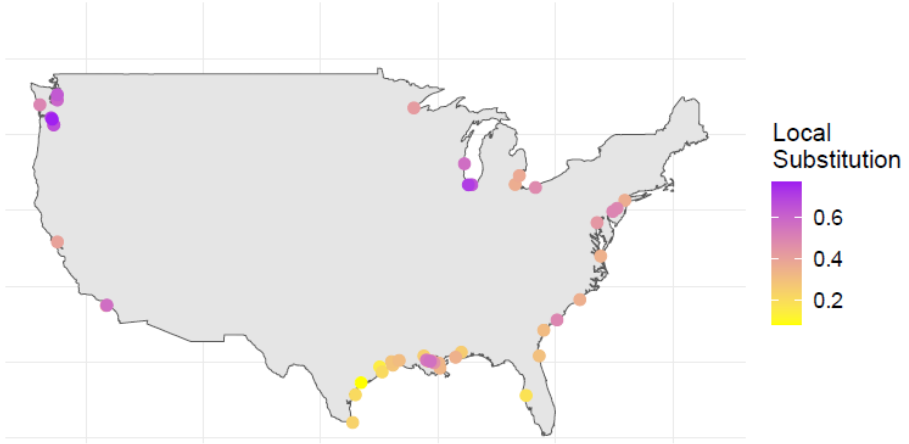

Figure 8: Local Substitution. For each port j this figure plots the fraction of shipments through j that is diverted away from j to a local port (i.e. to one of the 10 closest ports), conditional on shippers choosing to switch following a small increase in time at port.

5 Welfare Gains from Infrastructure Investment

Billions are spent every year from both public and private actors for investment in port infrastructure. What are the welfare gains of these investments? We now use our estimates for port technology and demand for port services to answer this question. In the next section, we compare these welfare gains to the infrastructure costs to obtain a measure of net returns.

We compute the aggregate welfare gains to an increase in the capacity of the US port system by one slot. In other words, a single port can now handle one additional ship; this is similar to the approach followed in Allen and Arkolakis (2022). To put this in context, this increase corresponds to approximately a 1% increase of the total US capacity. In practice, we increase K_j to $K_j + 1$ for a given port j and then we repeat this exercise for all J ports, one port at time. We use our model to simulate a long-run counterfactual time series of 30 years. We assume that along with K_j, port j's labor and cranes also increase so as to keep its (average) service time constant.[34] Note that since many small ports can handle only 1 or 2 ships at a time this increase in capacity can be substantial. We describe how these simulations are conducted in Appendix C.

Mechanisms We begin by outlining the economic mechanisms that are set in motion when the capacity of a port increases. We first describe the impact at the "treated" port, and then explore what happens to all other ports.

When capacity K_j increases at a **treated port** j, its wait time declines, all else equal. In turn, the reduction in congestion makes the port more appealing and attracts ships both from competing ports and from the outside good. It is important to note that the latter creates a net increase in total trade, unlike the ships' reallocation, which is a transfer of trade from other ports to port j.

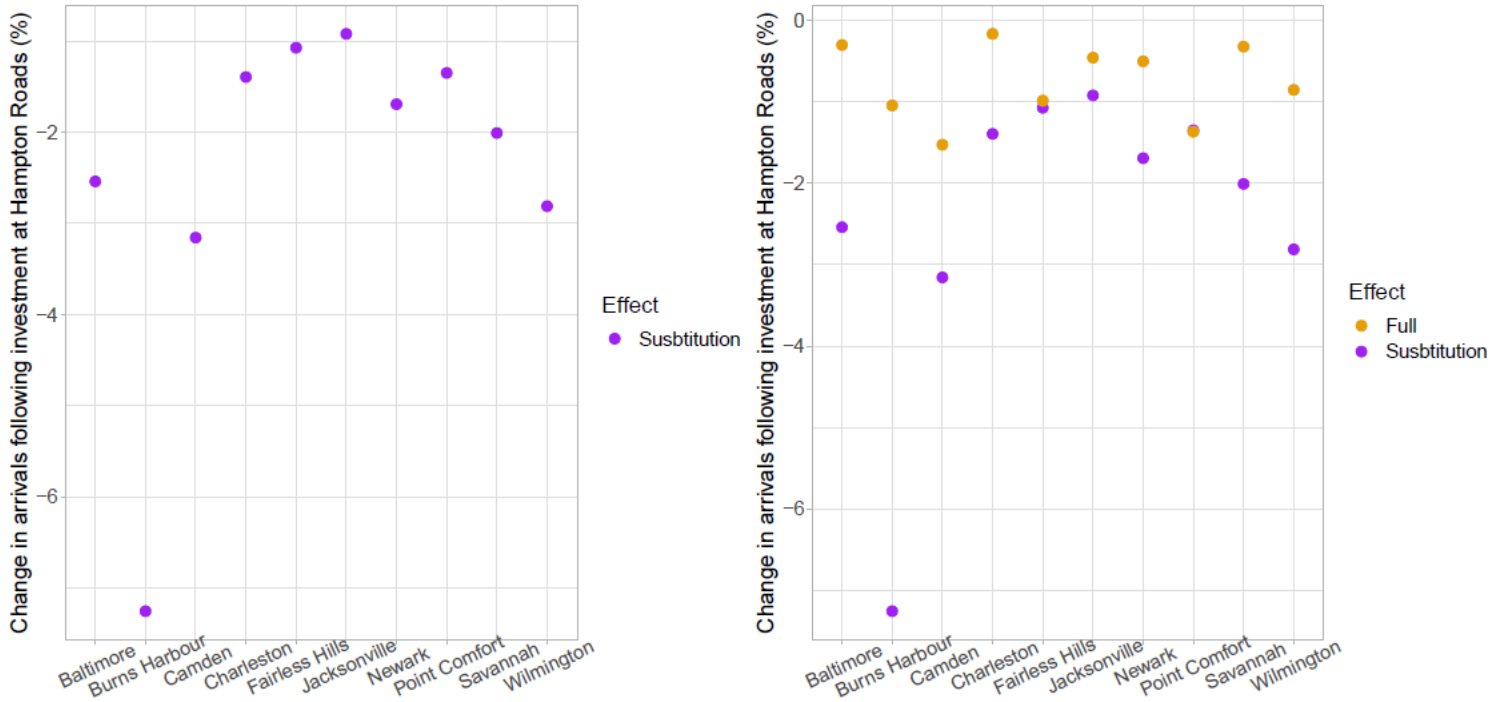

Figure 10: Substitution and full effect of increasing port capacity at Hampton Roads (VA) from K_j to $K_j + 1$, so that it can now handle one additional ship. The left panel plots the change in arrivals at ports other than Hampton Roads (VA), holding time at port at its original level there. The figure depicts the port's top 10 substitutes. The right panel plots the full effect on other ports' demand in response to an increase in infrastructure in Hampton Roads (VA) (i.e. not holding time at port constant at other ports).

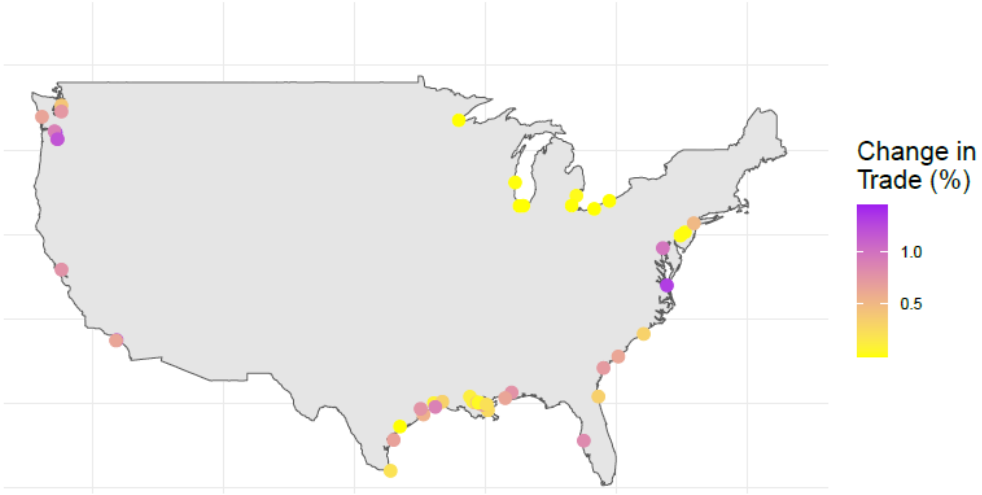

Figure 11: Trade impact of investment in infrastructure. In this figure, we increase port capacity of port j from K_j to $K_j + 1$, for all j one at a time, so that port j can now handle one additional ship. The figure depicts the total change in trade from increasing capacity in each port j, one at a time.

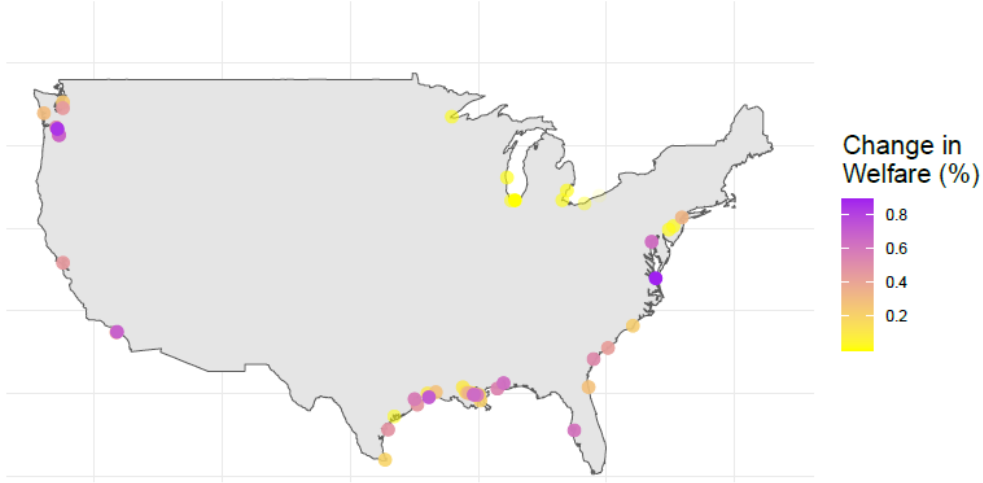

Figure 12: Welfare impact of investment in infrastructure. In this figure, we increase port capacity of port j from K_j to $K_j + 1$, for all j one at a time, so that port j can now handle one additional ship. The figure plots the total change in welfare from increasing capacity in each port j, one at a time.

and welfare can increase by more than 1% when investing in ports such as Hampton Roads (VA); but they are negligible when investing in other ports, such as the ones in the Great Lakes.[35] This suggests that targeting investment to the "right" port, as opposed to indiscriminate investment, can increase the returns to investment manyfold.

In addition, our estimates suggest that spillovers are large. In particular, if we only allow for the substitution effect, the aggregate increase in trade is 20% lower on average, and up to 39% lower for the ports with the highest gains.[36]

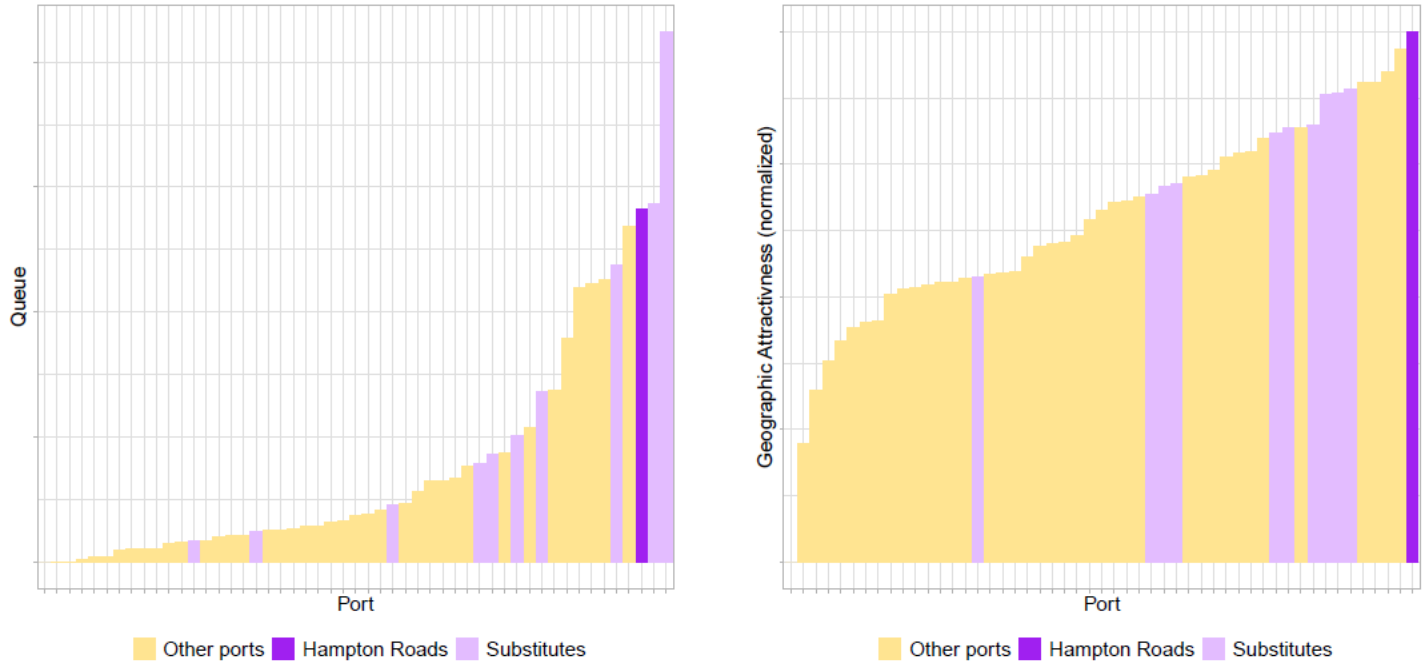

Figure 13: What makes a port generate high welfare gains? The left graph ranks ports by their queue. Hampton Roads (VA), the port with the highest returns, is fairly congested. In addition, its 10 closest substitutes are themselves also fairly congested. The right graph ranks ports by their geographic attractiveness, δ^{geo} . Hampton Roads is the most geographically attractive port and its 10 closest substitutes are themselves highly attractive ports.

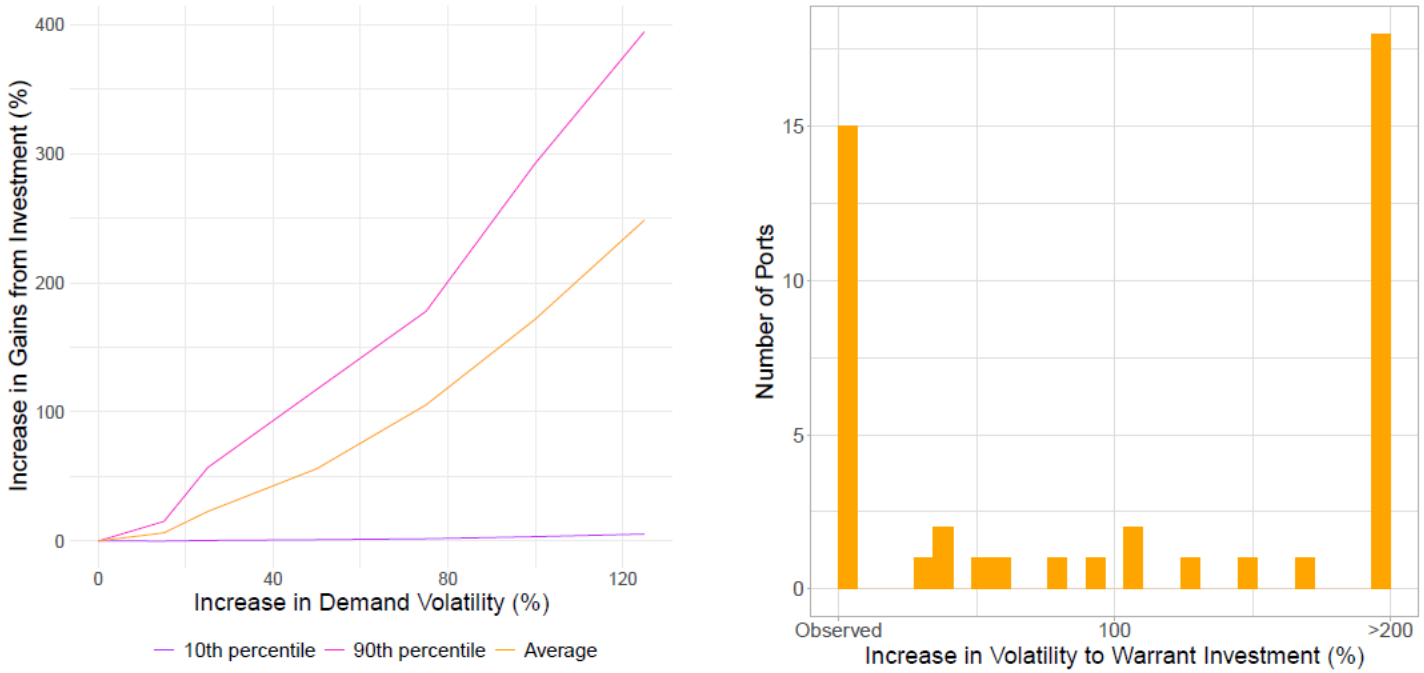

Figure 15: Rising macroeconomic volatility. The left panel plots the welfare gains from investment in the average, as well as the 10^{th} and 90^{th} percentile of US ports, for different percent increases of aggregate volatility. In particular, we repeat the long-run simulations under a mean-preserving spread of the aggregate demand process (i.e. we increase the standard deviation, σ_ϵ, of the aggregate demand γ_t). The right panel plots the histogram of the increase in aggregate volatility that turns net returns positive for different ports.

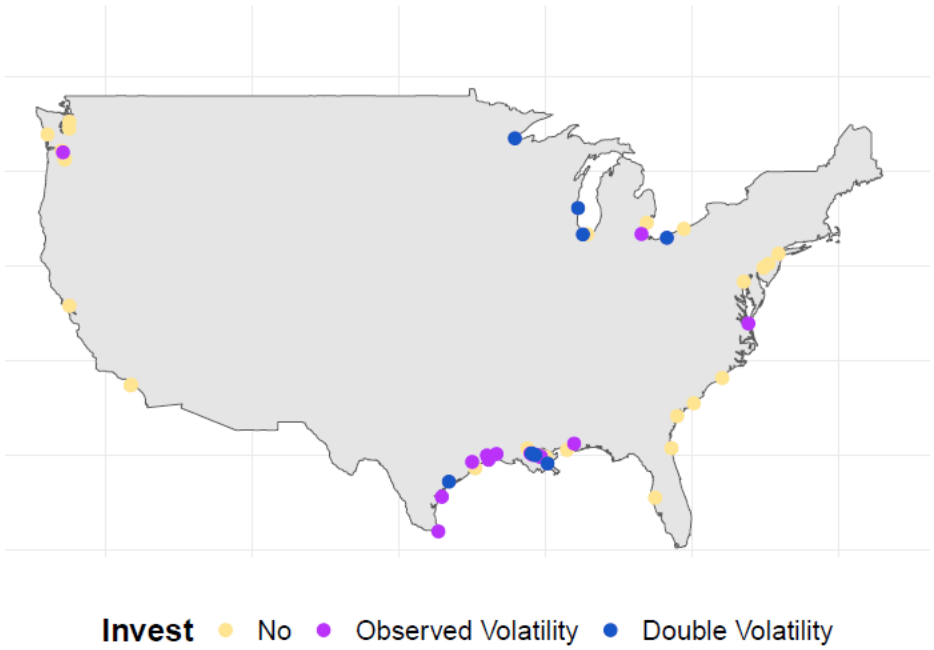

Figure 16: Rising macroeconomic volatility. This figure depicts the map of net returns: purple ports are the ports with positive net returns at the baseline level of volatility; blue ports are the ports whose returns turn positive when volatility doubles; and yellow ports are ports whose net returns remain negative.