NBER WORKING PAPER SERIES

MULTIMODAL TRANSPORT NETWORKS

Simon Fuchs
Woan Foong Wong

Working Paper 35065
http://www.nber.org/papers/w35065

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2026

We thank Treb Allen, Adina Ardelean, Costas Arkolakis, Bruce Blonigen, Matilde Bombardini, Mark Colas, Arnaud Costinot, Doireann Fitzgerald, Cecile Gaubert, Brian Greaney, David Hummels, Samuel Kortum, Nelson Lind, Yuhei Miyauchi, Andrii Parkhomenko, David Nagy, Esteban Rossi-Hansberg, Georg Schaur, Joseph Shapiro, Hugo E. Silva, Conor Walsh, and seminar participants at UC Berkeley Haas School of Business, Emory University, Hong Kong University, INSEAD, Kiel Institute, Minneapolis Federal Reserve Bank, Monash University, National University of Singapore, Purdue University, Seoul National University, Singapore Management University, UBC Sauder School of Business, University of Auckland, University of Chicago, University of Melbourne, University of Stavanger, University of Sydney, and University of Washington. We also appreciate the helpful feedback from workshop and conference participants including the Barcelona Summer Forum, NBER Conference on Economics of Transportation, Ron Jones Workshop (Rochester University), International Trade and Macro Workshop (St Louis Federal Reserve Bank), Trade Mini-Conference (UCLA), the 1st Annual World Bank Conference on Transport Economics, West Coast Trade Workshop (UC Davis), and West Coast Spatial Workshop (UC Berkeley). Benjamin Delgado and Philip Economides provided excellent research assistance. The views in this paper are solely the responsibility of the authors and should not necessarily be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System, any other person associated with the Federal Reserve System, or the National Bureau of Economic Research. All errors are our own. We are grateful to the NBER, DOT, and NSF for project support under the "Economics of Transportation in the 21st Century" Initiative (Award number SES-1559013, Project number 36363.03.08.00).

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2026 by Simon Fuchs and Woan Foong Wong. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Multimodal Transport Networks
Simon Fuchs and Woan Foong Wong
NBER Working Paper No. 35065
April 2026
JEL No. F11, R12, R42

ABSTRACT

Over half of distance-weighted U.S. freight is shipped using more than one transport mode. We examine how multimodal transport networks shape the economic and environmental impacts of infrastructure investments and disruptions. We develop a tractable spatial equilibrium model of multimodal routing with mode-specific congestion at intermodal terminals. We estimate a modal substitution elasticity using road and rail data, and a terminal congestion elasticity using vessel-positioning data. Calibrated to the U.S. freight network, the model identifies key bottlenecks and quantifies $.46-$1.85 billion in real GDP gains from intermodal terminal improvements, with additional environmental benefits from shifting away from carbon-intensive road transport. Ignoring mode-specific congestion overstates welfare gains from highway improvements by 85%, while ignoring multimodal flexibility understates them by 22%. Losing rail network access is estimated to reduce real GDP by $230 billion.

Simon Fuchs
Federal Reserve Bank of Atlanta
sfuchs.de@gmail.com

Woan Foong Wong
University of Oregon
Economics Department
and NBER
wongwoanfoong@gmail.com

Freight transportation is inherently multimodal: over half of U.S. freight ton-miles is transported using two or more modes over an integrated network of roads, railways, barges, connected by ports and inland terminals (Department of Transportation). Recent global disruptions, from pandemic-related port congestion to the Panama Canal drought and U.S. rail strike threats, have highlighted how localized vulnerabilities can propagate throughout the broader multimodal transport network to significantly impact trade costs, supply chains, and welfare (World Bank, 2023; National Economic Council, 2024). Furthermore, since transportation is the largest source of direct U.S. greenhouse gas emissions (EPA, 2024) and transport modes differ substantially in their emissions intensity——trucks emit eight times more CO_2 per ton-mile than rail (CBO, 2022)——improvements and disruptions to the transport network can have important implications for climate change.

In this paper, we study how the multimodal structure of transport networks shapes the economic and environmental impacts of infrastructure investments and disruptions. We do so by developing a spatial equilibrium model of freight routing across multiple modes with mode-specific congestion at intermodal terminals, which are key transition points where freight shifts between modes. To identify how freight reallocates across modes, we estimate an elasticity of modal substitution using confidential U.S. rail shipment data combined with truck traffic data. To capture congestion effects, we use minute-level vessel positioning data to estimate congestion elasticities at intermodal port terminals. Incorporating these features allows us to quantify the system-wide effects of infrastructure investments and major disruptions taking into account the multimodal nature of transport networks. Relative to our approach, we show that single-mode frameworks can underestimate the welfare effects of infrastructure investments.

We account for the complexities of modeling multimodal transport networks within a general equilibrium setting by incorporating recent advances in the operations research literature (Daly and Bierlaire, 2006). Specifically, we develop a model that recursively specifies optimal choices for sourcing, routes, *and* modes. There are three main benefits to our framework. First, the model embeds the choice of transport modes within an optimal route choice model, with a separate elasticity of substitution across modes. This structure can flexibly nest canonical single-mode routing models as a special case. Second, it can include a rich set of mode-specific congestion patterns, including capacity constraints at intermodal terminals. Third, the model can derive closed-form expressions for equilibrium transport costs, traffic flows, and market access terms, which depend endogenously on mode and route choices as well as the topology of the transport network.

We formally demonstrate that mode-specific changes within the multimodal transport network generate both direct and indirect traffic adjustment effects that propagate through the entire network. First, a reduction in the cost of one mode on link kl diverts traffic toward that mode and away from competing modes on the same link, generating a direct *modal substitution* effect. An improvement in road transport on a given link, for example, shifts traffic on that link away from rail and toward road. Second, because the cheaper mode lowers the composite transport cost of the link, it raises the attractiveness of

routes and locations that use it and increases transport demand more broadly across the network; this is an indirect *modal complementarity* effect. The strength of this complementarity channel is greatest for links that are central to network connectivity and have few close alternatives, and its ultimate importance depends on how the induced traffic interacts with congestion across links and at terminals.

We then connect these traffic adjustments to welfare. Applying the inverse function theorem to our equilibrium system we derive *closed-form decompositions* of the welfare elasticity associated with a mode-specific link improvement which highlights two distinct margins. First, the *reallocation wedges* capture how the welfare of this improvement changes when traffic is allowed to reallocate across routes and modes. They are positive when it reinforces the improvement by directing traffic toward especially valuable uses of the affected link, and negative when the additional flexibility mainly creates bypasses that weaken its importance. Second, the *congestion wedges* measure how much of the welfare gains is absorbed once the induced traffic loads onto congested links and terminals within the network. Taken together, these wedges show whether a link-level improvement creates value because it reallocates traffic in useful ways, or whether part of that gain is dissipated because the additional traffic increases congestion elsewhere in the system. Using this decomposition, we can isolate and quantify the contributions of the main ingredients in our model by sequentially shutting down components of these two margins.

Our analysis relies on two central parameters: the elasticity of substitution across modes and impact of congestion at intermodal terminals. We estimate the elasticity of modal substitution by studying how traffic shifts between modes in response to cost changes in a one mode, specifically how relative rail and road traffic flows respond to a decrease in road costs. By matching confidential waybill rail traffic data to truck traffic in cities, our approach combines the mode-specific cost change research design by Duranton and Turner (2011), which approximates reductions in road cost with increases in interstate highways, with the analysis on relative traffic shift between modes as in Hummels and Schaur (2013) and Lugovskyy, Skiba and Terner (2022). We find that a 1 percent increase in interstate highways results in a 0.9-1.2 percent decrease in relative rail to truck traffic use. Using the most conservative specification, we estimate the elasticity of modal substitution to be 1.099.

Next, we evaluate the impact of congestion at intermodal terminals using minute-level vessel positioning data, which allows us to estimate how responsive ship dwell times (time spent loading and unloading at port) are to overall port traffic at the time of arrival. To identify the causal effect of overall port traffic on ship dwell times, we employ a shift-share instrument that serves as a demand shifter for port services. We find that a 1 percent increase in port traffic increases ship dwell times by 0.24-0.26 percent, which corresponds to an intermodal terminal congestion elasticity of 0.096. We demonstrate the broader applicability of this elasticity to the entire multimodal network, including rail terminals, by highlighting a significant positive correlation between port traffic and railcar dwell times.

To conduct counterfactuals, we calibrate our framework to the U.S. transportation network using high-resolution geo-spatial data on roads, railways, waterways, and the location of intermodal terminals

(both inland and ports). Our first application quantifies the welfare impact of improving the integration of the multimodal transport network by running 228 counterfactuals, each simulating a reduction in mode-switching cost at a different intermodal terminal in the U.S.. We find that terminals generating the largest gains are centrally located in the interior of the U.S., for example in Chicago and Atlanta, highlighting the role of the multimodal network transporting goods from coastal to interior regions. Terminals in major urban centers, like Los Angeles and Houston, also rank highly. Improving these central terminals would generate an aggregate welfare gain equivalent to an additional $463-3,851 million of GDP (in 2012 USD). Congestion significantly affects these gains: welfare benefits are 2.7 times higher on average without congestion. These investments also result in modal substitution, shifting traffic use from road to rail and barges. Since trucks generate more greenhouse gases relative to trains, these substitution effects result in environmental benefits valued at $17-214 million. These findings align with the goals of the inter-agency U.S. National Blueprint for Transportation Decarbonatization, which emphasizes improving multimodal freight transport (DOE, DOT, EPA, and HUD, 2024).

Building on the theoretical decomposition earlier, we quantify the importance of four main channels driving welfare gains to transportation infrastructure improvements in our model: (1) road and terminal congestion, (2) productivity and amenity externalities, (3) flexibility across multiple transport modes, and (4) endogenous route choice. We implement this decomposition in a link-level U.S. highway improvement setting, evaluating 704 counterfactuals that improve each highway link individually for each specification. Our decomposition shows that standard approaches can substantially misestimate the welfare effects of transportation improvements. Ignoring congestion leads to an 85 percent overestimate of welfare gains, while abstracting from productivity and amenity externalities overstates gains by 57 percent. Conversely, shutting down multimodal flexibility—effectively reducing the network to a single mode—understates welfare gains by 22 percent, and fixing routes understates them by a similar magnitude, which suggests most of the rerouting response occurs across modes rather than within modes.

Additionally, we quantify the welfare and environmental impacts of three policy scenarios. First, we evaluate the value of the U.S. multimodal network by exploring the effects of the 2022 potential rail strike which would have shut down the entire rail transport system (Kanno-Youngs and Cochrane, 2022). While existing literature has quantified the significance of the road and highway transport network, there has been less attention on the value of the rail network apart from historical analyses. We find that losing access to the rail transport network would decrease U.S. GDP by approximately $230 billion (in 2012 USD), inclusive of the mitigating effects of modal substitution towards roads and waterways, as well as the broader general equilibrium effects of reduced market access due to the loss of rail. This welfare loss is exacerbated by congestion effects and would generate additional environmental damage as transport shifts towards road transport, adding an estimated social cost of $12 billion.

Second, we consider the consequences of repealing the Jones Act of 1920, an active U.S. trade regulation that requires all ships transporting goods and passengers between U.S. ports to be U.S.-built

and majority-owned and crewed by U.S. citizens. By repealing this act and allowing cheaper foreign ships to operate on U.S. domestic waterways, we estimate a welfare gain equivalent to increasing U.S. GDP by about $3.2 billion (in 2012 USD). Our finding falls on the lower end of existing government estimates, likely due to our focus on the continental U.S. and because the multimodal network has had a century to adjust away from domestic waterways under the Jones Act. Had waterborne transport been less expensive, the dense rail and road corridors that now line the U.S. coastlines might well have been built further inland instead. The repeal of the Jones Act would also allow for less reliance on trucks and rail, which would result in an environmental benefit of $0.2 billion since waterway transport is greener.

Our third scenario investigates the impact of the Panama Canal disruptions due to decreased access from drought conditions (Rojanasakul, 2024). The resulting welfare loss is equivalent to a $2.7 billion decrease in U.S. GDP (in 2012 USD). To our knowledge, this is the first quantification of the Panama Canal disruption which accounts for both modal and route substitutions. We find environmental losses of about $0.5 billion due to the substitution towards both trucks and rail. Across all three scenarios, we show that congestion—at terminals and for both terminals and roads—has a compounding effect on welfare. For the Panama Canal disruptions, this is particularly pronounced: without terminal congestion, the welfare losses increase by more than 186 percent. This large increase is driven by modal shifts from water to rail and road, and the crucial role of intermodal terminals in facilitating this shift. Removing both terminal and road congestion increases the effect by a further 35 percent.

Related Literature Our paper is complementary to and builds on a number of different strands of research. First, this paper contributes to a rapidly expanding literature incorporating realistic transportation networks into quantitative spatial equilibrium models (see Donaldson (2025) and Allen and Arkolakis (2025) for recent surveys). Within that literature, there have been multiple efforts to merge the disaggregated network structure of transportation infrastructure with a general equilibrium economic geography model (Donaldson and Hornbeck, 2016; Fajgelbaum and Schaal, 2020; Allen and Arkolakis, 2022; Fajgelbaum et al., 2023). Most relevant to our work, Allen and Arkolakis (2022) (hereafter abbreviated to AA2022) propose a tractable way of incorporating the optimal routing choice into a spatial equilibrium model to examine the general equilibrium implications of infrastructure improvements. Our paper builds on this seminal work and this literature by developing a model that recursively specifies optimal choices for sourcing, routes, *and* modes. Our approach allows us to embed mode choice within an optimal route choice model with a separate elasticity of substitution across modes, and identify how bottlenecks emerge from the multimodal transport network due to mode-specific congestion and limited substitution across modes.[1]

Additionally, this literature has often focused on one mode of transport—road, maritime, *or* rail (Coşar

[1]Our paper is also complementary to the literature on optimal transport infrastructure policies (Santamaria, 2020; Almagro et al., 2024; Fajgelbaum and Schaal, 2020), while we do not solve for optimal policy directly, we identify how bottlenecks emerge within multimodal networks which can help inform the targeting of infrastructure investments.

and Demir, 2018; Brancaccio, Kalouptsidi and Papageorgiou, 2020; Heiland et al., 2019; Ganapati, Wong and Ziv, 2021; Degiovanni and Yang, 2023; Dunn and Leibovici, 2023; Gibbons, Heblich and Pinchbeck, 2024; Asturias, 2020; Economides, 2024; Do et al., 2025). More recent and complementary work has expanded this analysis to more modes of transport. Focusing on roads, Fan, Lu and Luo (2019) study how domestic road and highways affect transport costs and welfare in China respectively.[2] Jaworski, Kitchens and Nigai (2023) quantifies the value of U.S. highways allowing for modal substitution for transportation across states, not within. Our paper adds to this literature by studying the general equilibrium analysis of the U.S. multimodal transport system—roads, barges, *and* rail—as well as intermodal terminals which allow for switching across modes. We estimate an elasticity of modal substitution that allows us to pin down how freight reallocates across rail and truck.[3] Combined with our model, we can quantify the economic and environmental impacts of infrastructure investments and real-world policy scenarios while accounting for substitution and complementarity within the multimodal network, and show that models limited to a single transport mode can understate these impacts.

More recent work has focused on the importance of ports and the gains from port development and efficiency. Brooks, Gendron-Carrier and Rua (2018) examines how the adoption of container technology at ports affects local activity, while Ducruet et al. (2024) identify how this technology adoption can crowd out economic activity by increasing land rents. Bonadio (2021) investigates how improvements in roads and ports have different distributional consequences locally and abroad. Focusing on dry bulk shipping, Brancaccio, Kalouptsidi and Papageorgiou (2024) evaluates the returns to infrastructure investment at ports in the presence of disruptions and congestion. Our paper adds to this literature by studying intermodal terminals, both rail terminals and coastal ports, in the context of the multimodal transport network, showing that they are not just important for market access internationally but also domestically. Furthermore, we estimate an elasticity of congestion at intermodal terminals, and show that congestion—both at intermodal terminals and on roads—can substantially compound the effects of network improvements and disruptions: we show that abstracting from congestion can overstate both the welfare gains from both terminal and highway-link improvements .

Our paper is also related to a long-standing theoretical and empirical literature in transportation on route and mode choice (McFadden, Winston and Boersch-Supan, 1986; Rich, Kveiborg and Hansen, 2011; Beuthe, Jourquin and Urbain, 2014; Winston, 1981). The state-of-the-art in transportation studies solves high-dimensional traffic assignment problems algorithmically (see Notteboom (2018) for a recent survey). We employ similar tools, specifically the *stochastic user equilibrium* where routes and modes are chosen subject to a stochastic perception error. However, we go beyond the extant literature by fully

[2]Additionally, Fan and Luo (2020) is a note which characterizes bilateral transport costs and their elasticities with respect to transshipment costs.

[3]Our estimate is complementary to work on modal substitution between ocean and air freight (Harrigan, 2010; Hummels and Schaur, 2013; Lugovskyy, Skiba and Terner, 2022; Tolva, 2025) and between bulk and container shipping within ocean freight (Coşar and Demir, 2018). Compared with ocean–air studies, our effects are smaller, likely because the high cost of air shipping implies a larger elasticity of substitution.

embedding the stochastic user equilibrium into a spatial general equilibrium framework, where input and output markets across space clear and factor and output prices are endogenously determined. This allows us to quantify the welfare effects of infrastructure improvements and disruptions within the multimodal transport network, capturing both modal substitution and general equilibrium complementarity effects.

Finally, our paper is related to the recent literature on the environmental impacts of transportation. Most of this literature focuses on the link between international trade and greenhouse gas emissions via transportation, and changes to this relationship in response to environmental regulations (Shapiro, 2016; Mundaca, Strand and Young, 2021; Lugovskyy, Skiba and Terner, 2022; Ludwig, 2025) or trade policies (Cristea et al., 2013). We instead highlight how infrastructure investments and disruptions can have environmental consequences via the multimodal transport network. While some of these papers have found compositional shifts in transport mode use due to regulation and trade policy changes,[4] our quantitative general equilibrium framework allows us to distinguish between the substitution and complementarity effects. Given that transportation accounts for a quarter of global emissions (United Nations, 2021), our framework allows us to investigate how both of these effects contribute towards the environmental impacts of transportation infrastructure improvements and disruptions.

The rest of the paper is structured as follows. Section 2 describes our data and the U.S. multimodal transport network. Section 3 presents the multimodal spatial equilibrium routing model, and Section 4 explains the estimation of modal substitution and congestion elasticities. Section 5 describes the calibration of the model to the U.S. multimodal network and provides validation checks of our framework. In Section 6, we conduct counterfactual analyses to quantify the welfare and environmental impacts of infrastructure investments and policy scenarios, and decompose the results to highlight the model's key channels. Section 7 concludes.

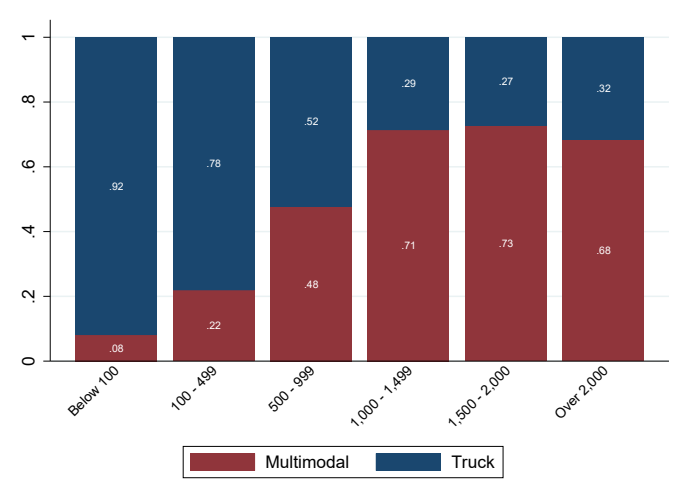
2 Data: U.S. Domestic Freight Transportation

In this section, we introduce our data sources and provide an overview of the U.S. domestic transport system. Figure 1 plots the share of U.S. freight moved by different modes across various distances. Trucks primarily handle freight for shorter distances under 100 miles, accounting for 92% by weight (similarly by value, Figure A.1), highlighting the role of the denser road network in facilitating transportation at the start and end of the movement of goods. This is commonly known as the first and last mile in freight transportation (Rodrigue, 2020; Ranieri et al., 2018). Freight moved over longer distances is increasingly transported using multiple modes. For freight transported over 500 miles, multimodal transport accounts for almost half by weight and for longer distances above 1,000 miles this share increases to more than

[4] Cristea et al. (2013) finds that trade liberalization between countries will increase trade from more distant partner countries, resulting in a proportional increase in air transport use and greenhouse gas emission. Lugovskyy, Skiba and Terner (2022) finds that environmental regulations capping CO_2 emissions from maritime shipping will substitute demand towards air transport, increasing total transport-related CO_2 emissions.

two-thirds.[5] On average, over 65% of U.S. freight shipped more than 500 miles uses multiple transport modes. For context, the road distance between Los Angeles and Chicago is about 2000 miles. Overall, more than half of U.S. freight in ton-miles is transported using more than one mode.

Figure 1. U.S. Transport Mode Shares by Distance



Notes: This figure plots the observed weight share of cargo transported by different modes across various distances in 2018. Multimodal indicates cargo movement that involves more than one mode, including truck, rail, and waterways. See Figure A.1 for the observed value share of cargo across various distances. For freight transported over 1,500 miles, multimodal transport accounts for more than half by value. Both value and weight shares are similar when including multimodal shipments by air. Source: Freight Analysis Framework, U.S. Department of Transportation, and authors' calculations.

2.1 Rail, Road, Waterways, Ports, and Intermodal Terminals

Rail Geography Network and Traffic. We use detailed geo-spatial information from the Topologically Integrated Geographic Encoding and Referencing Database (TIGER, Census Bureau) to construct a graph representation of the U.S. intermodal rail network. We subset the original network to segments owned by the largest Class I carriers and that are compatible with multimodal transport.[6]

For rail traffic data, we obtained access to confidential carload waybills data from the Surface Transportation Board (see Section B.1.1 for further details). The waybill data captures the origin, destination, and interchange rail stations for freight cargo transported through the rail network. This detailed geographical information for rail cargo journeys allows us to capture their routes. Additionally, the data set contains information on number of car loads, weight, and intermodality. Intermodality indicates if rail freight movement involved other transport modes, which is almost entirely containers. The proportion of freight transported over multiple modes has increased dramatically over time: intermodal

[5] Multimodal transport accounts for more than half by value for freight transported over 1,500 miles (Figure A.1).

[6] Class I railroads are the largest carriers operating on the U.S. railroad system, originally defined in 1992 to be carriers above $250m dollars of revenue. Since adjusting for inflation, the cutoff in 2021 stood at approximately $943m. The seven class I carriers currently in operation comprise the large majority of the domestic rail freight market.

container rail traffic is the fastest growing rail traffic segment, having grown by more than 5 times since 1984 (Figure A.2). We restrict our rail traffic sample to intermodal freight transported by Class I carriers.

Figure 2a presents the U.S. intermodal rail traffic flows. Thicker lines, indicating higher traffic flows, link the U.S. West and East coastal regions to interior locations like Chicago, Dallas, and Atlanta. Examples of such routes include Los Angeles to Chicago, New York City to Chicago, and Los Angeles to Dallas. Consistent with Figure 1, the rail network serves to transport freight over long distances.

Road Geography Network and Traffic. We follow AA2022 in constructing road geo-spatial and traffic flows data (see Section B.1.2 for details). For road traffic data, we use the average annual daily traffic from the 2012 Highway Performance Monitoring System (HPMS) dataset by the Federal Highway Administration. Flows are allocated to individual links by constructing a length-weighted average of the annual daily traffic. Figure 2b presents a graph representation of the interstate highway system with thicker lines indicating higher road traffic flows. The highway flow patterns are quite different from rail (Figure 2a). Here, heavier highway traffic connects large densely populated cities that are either on the coast or in the interior regions, like Los Angeles to San Diego, Boston to Philadelphia, and the surrounding areas around Chicago.

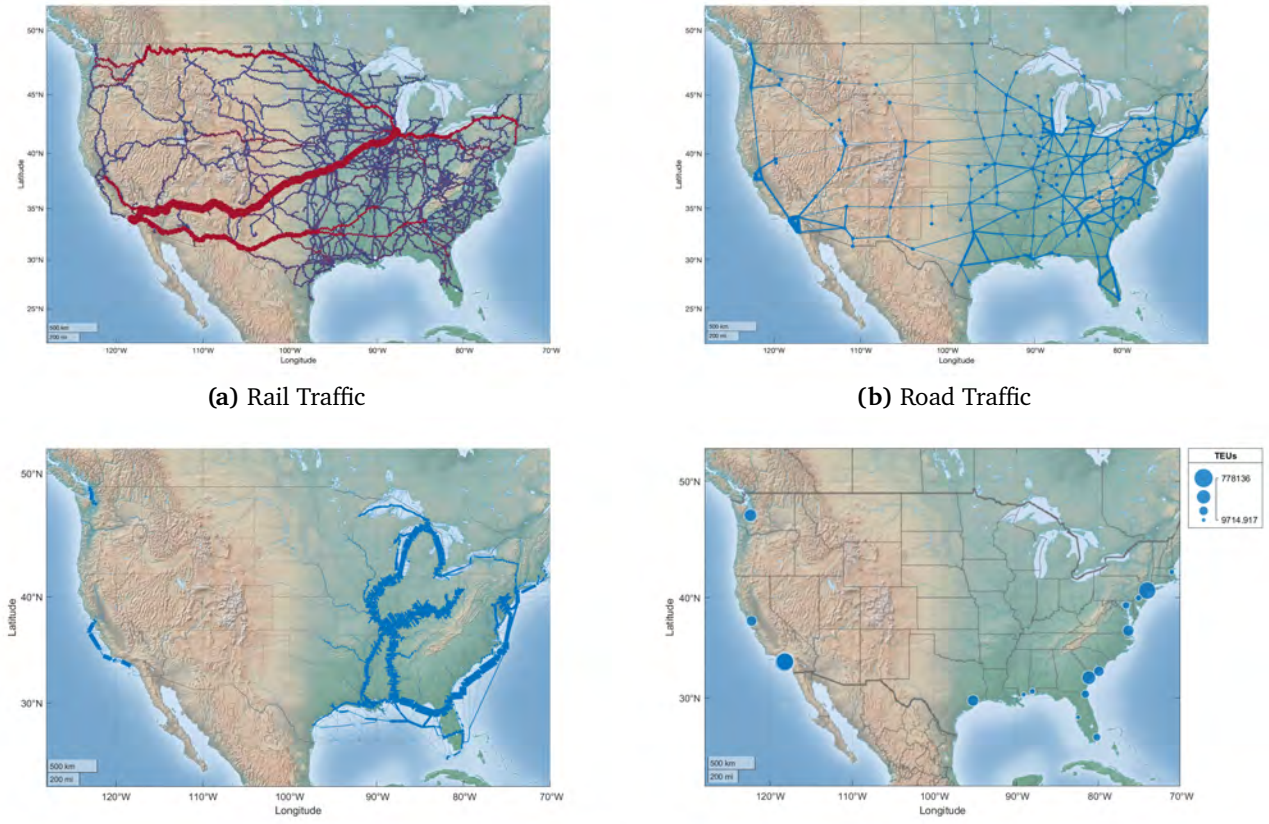
Waterway Traffic. We next capture goods transported via U.S. waterways by bringing in waterborne traffic data and location of inland ports from the U.S. Army Corps of Engineers (USACE). We restrict our sample to manufactured goods which are primarily transported by barges. Figure 2c shows that most barge traffic is concentrated on the Great Lakes, Illinois River, and Mississippi River.

Port Location and Traffic. We include the top U.S. container ports as these are important coastal intermodal terminals facilitating the transfer and distribution of goods between ships and trucks or rail (see Section B.1.3 for details). Geospatial data on port location comes from the USACE. Container volumes at these ports come from the Port Performance Freight Statistics Program maintained by the Department of Transportation (DOT). Figure 2d visualizes the coastal ports in a geographic bubble map, where the largest container ports include Los Angeles/Long Beach and New York (Newark). Between the coastal and inland ports, we capture 98 percent of total U.S. container volumes.

Intermodal Terminals. Intermodal terminals also include inland transfer points for goods moving between rail, road, and waterways. We capture the location of these intermodal terminals using the National Transportation Atlas Database (NTAD) maintained by the Department of Transportation (DOT).

U.S. Multimodal Transport Network. To construct the U.S. multimodal transport network, we first combine all the geo-spatial information for each of the rail, road, and waterway networks. We then add the locations of intermodal terminals, including coastal and inland ports, as well as inland intermodal terminals. Figure 3 illustrates the resulting graph representation of the U.S. multimodal transport

Figure 2. Traffic along the U.S. Domestic Freight Transport System



(a) Rail Traffic

(b) Road Traffic

(c) Waterway Traffic

(d) Port Traffic

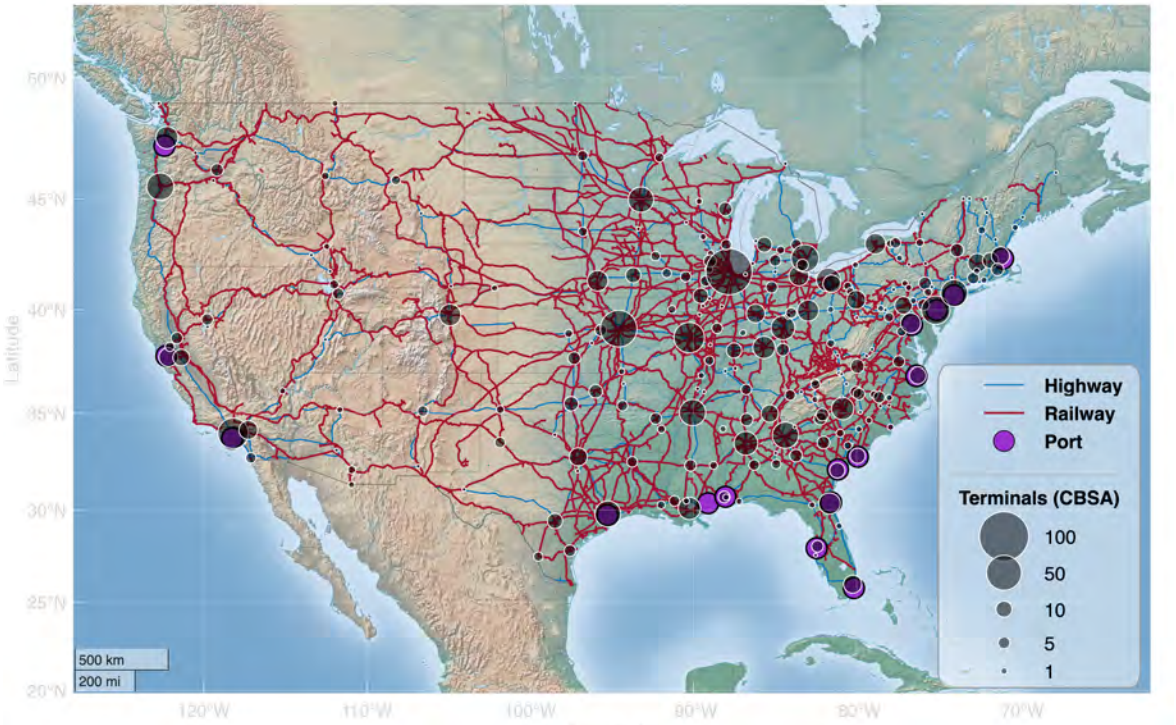
Notes: Panel (a) presents U.S. domestic rail traffic for Class I carriers (the largest rail carriers) conditional on intermodal capability. Shortest routes are imputed between origin, interchange stations, and destination to assign total tonnage to individual rail segments along network. Thicker lines indicate higher traffic flows. Panel (b) presents U.S. domestic road traffic along the graph representation of the interstate highway system. Thicker lines indicate higher traffic flows. Panel (c) presents U.S. domestic waterborne traffic for manufactured goods, where the shortest routes are imputed between origin and destination to assign total tonnage to individual segments of the domestic water network. Thicker lines indicate higher traffic flows. Panel (d) presents the container traffic volume at international ports. Larger circles indicate higher container volumes. Sources: Authors' calculations, Confidential Carload Waybill, Surface Transportation Board; Highway Performance Monitoring System, the Federal Highway Administration; Waterborne Commerce statistics, U.S. Army Corps of Engineers (USACE); and Port Performance Freight Statistics Program, Department of Transportation (DOT).

network consists of 288 nodes and 704 edges. The nodes are either core-based statistical areas (CBSAs) or endpoints and intersections in the network.

2.2 Automatic Identification System (AIS) Vessel Traffic Data

We utilize automatic identification system (AIS) vessel traffic data from Marine Cadastre, a joint initiative between the Bureau of Ocean Energy Management and the National Oceanic and Atmospheric Administration. This data captures vessel location in U.S. waters at 1-minute intervals using 200 land-based receiving stations. We observe the vessel's unique identifying International Maritime Organization Vessel number (IMO) as well as longitude and latitude location down to the minute, speed, and navigation status. The vessel's navigation status captures whether the vessel is being propelled (under way using engine), or moored (held in position at a pier). Using information on the ship's speed and navigation status, we define a ship's dwell time as the time it spends moored at a pier at zero speed. This is a

Figure 3. U.S. Multimodal Transportation Network



Notes: This figure shows the combined U.S. multimodal freight network. The original GIS information comes from the U.S. Census Bureau's Topologically Integrated Geographic Encoding and Referencing (TIGER) Database. Red lines indicate the Class I multimodal railroad network. Blue lines indicate the interstate highway system (IHS). Black diamonds indicate freight terminals that are owned by Class I operators allow for road-to-rail or rail-to-road intermodal movements. Purple circles indicate the top 18 ports.

conservative measure of ship dwell time at ports because (1) ships spend time navigating within the port area as they prepare to moor at a pier and (2) ships can also end up waiting outside of the port area at anchor before navigating to the port (Schmidt, 2021).

To match these ships to the ports where they are located, we use the USACE Port Statistical Area shapefiles and match these ships to the top 30 U.S. container ports. These port polygon areas also allow us to calculate the total amount of time a ship spends within the port region on top of the time it spends moored at a dock. Additionally, in order to identify the cargo capacity of these ships and their containership status, we match these ships to the USACE Port Entrance and Clearance dataset using their identifying information and when they are at these ports. See Section B.1.4 for examples illustrating how we observe these ships and the duration of their stay at ports. The ship cargo capacity measures the volume of the ship that can be used for loading cargo (net tonnage of a ship). This cargo capacity measure for each ship will contribute to our port traffic measure at each port every day.

Port Traffic Our measure of port traffic is defined as the sum of the net tonnage of each ship moored at the port each day, multiplied by the percent of the day they spend at the port. Crucially, this measure includes ships that arrived on a prior day but remain moored at port. As an example, a ship that remained moored at port all day without exiting would contribute 100% of its net tonnage to the measure of port traffic. If that ship left at any point during that day, its net tonnage contribution would be instead

determined by the amount of time it spent moored at port that day. With this daily port traffic measure, we calculate moving averages of the port-level traffic for varying time windows (3, 7, 14, 21, and 28 days). We present the 28-day moving average results in the main text and provide the remainder in the Online Appendix B.4.

Ship Dwell Times Summary Statistics Our matched dataset has 3,755 unique vessels with 1,444 containerships. The top 30 ports in our dataset account for around 95% of all U.S. container trade by value and 98% by volume annually. For illustrative purposes, Figure 4a aggregates our minute- and ship-level dataset and plots the average of monthly containership dwell times across the top 30 U.S. ports from June 2015 to December 2021. The average dwell time over this period is around 33.3 hours per ship with a standard deviation of 5 hours.[7] However, as seen in Figure 4a, there is a significant increase in ship dwell times after 2021, averaging 42.8 hours.

Additionally, there is considerable variation in ship dwell times across different ports and time periods. Figure 4b compares the average dwell times at the ports of Long Beach and Los Angeles (LA) to the overall average from Figure 4a. The dwell times at both ports, 73.6 hours and 82.1 hours respectively, are more than twice the overall average. After 2021, dwell times at these ports surged even more dramatically, increasing to over 4 days at Long Beach (104 hours) and almost 6 days at LA (136 hours). This substantial variation highlights the growing congestion and delays at key intermodal terminals and the underlying interaction between ships and ports contributing to this congestion.

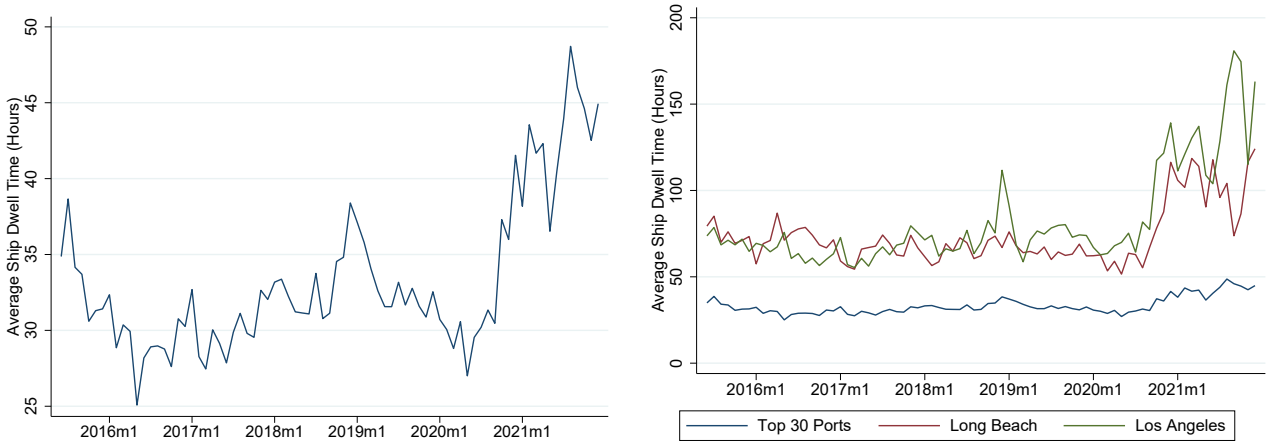
3 Economic Geography Model with Multimodal Routing

In this section, we develop a tractable model of routing across multiple transport modes with mode-specific congestion, especially at intermodal terminals, and embed it in an otherwise standard economic geography model with domestic trade between discrete locations and freely mobile labor reallocation across locations (Allen and Arkolakis, 2014; Redding, 2016). Two technical challenges arise in this environment. First, the number of feasible routes can be very large even in low-dimensional graphs, and the presence of multiple transport modes further compounds this path-enumeration problem. This curse of dimensionality makes brute-force solutions impractical. Second, existing route-choice approaches based on invertible matrix expressions, such as AA2022, do not readily accommodate a separate elasticity of substitution across modes. We address both issues by formulating route choice recursively and nesting mode choice at each segment along the route.[8]

[7] Our averages are lower than the estimates in Brancaccio, Kalouptsidi and Papageorgiou (2024) in part due to their inclusion of wait time at anchorage and their focus on bulk ships.

[8] One alternative method to introduce mode choice is to simply allow for mode-specific subnetworks. This naturally gives rise to a setting where mode choice is then *implied* by route choice, which represents a direct extension of AA2022 to transport mode choice (see Appendix A.5 for details). A previous version of this paper presents such a model. The limitation of this modeling approach is that it assumes that the elasticity of modal substitution must be identical to the elasticity of route

Figure 4. Containership Dwell Times at Port



(a) Overall Average at top 30 Ports **(b)** Long Beach and Los Angeles vs Overall Average

Notes: Panel (a) plots the overall average of monthly containership dwell times across the top 30 U.S. ports from June 2015 to December 2021. Panel (b) plots the monthly containership dwell times at the top two largest U.S. ports, Long Beach and Los Angeles (red and green lines respectively), and compares their dwell times to the same Panel (a) overall average (blue line). Both panels are weighted by ship net tonnage. Sources: Authors calculations, Marine Cadastre, Bureau of Ocean Energy Management and the National Oceanic and Atmospheric Administration; International Maritime Organization; and Port Statistical Area, U.S. Army Corps of Engineers (USACE).

The main idea is to recast the equilibrium problem in terms of local link-level objects rather than the universe of full paths. This makes it possible to introduce multimodal routing, mode-specific congestion, and terminal bottlenecks into a spatial equilibrium framework while retaining closed-form transport and traffic objects that can be taken to the data. It also provides a convenient way to separate the two distinct substitution margins in the model: substitution across routes and sources, and substitution across transport modes.

We proceed as follows. Subsection 3.1 defines the multi-layered graph that represents the multimodal transport network, and specifies consumption and production. Second, Subsection 3.2 introduces the recursive routing choice. Subsection 3.3 then expands on this to include the combined routing and sourcing choice, and resulting equilibrium. Next, Subsection 3.4 presents the nested transport mode choice while Subsection 3.5 introduces congestion and the gravity equations for traffic flows. Finally, to provide further intuition for the model mechanisms, Subsection 3.6 characterizes how changes in transport costs propagate through the network, shaping traffic and welfare outcomes.

3.1 Setup

Geography and transport Let $\mathscr{G} \equiv (\mathscr{N}, \mathscr{L})$ be a multigraph representing a multimodal transport network where $\mathscr{N}$ and $\mathscr{L}$ are the set of nodes and links respectively. We define the set of successor nodes $\mathscr{F}(i)$ and the set of predecessor nodes $\mathscr{B}(i)$ for each node $i \in \mathscr{N}$. Furthermore, let $\mathscr{G}_m \equiv (\mathscr{N}_m, \mathscr{L}_m)$ be the subgraph representing the transport network for mode m, where $\mathscr{N}_m \subseteq \mathscr{N}$ and $\mathscr{L}_m \subseteq \mathscr{L}$ are the mode-specific nodes and links respectively. We also define the mode-specific successor and predecessor nodes $\mathscr{F}_m(i)$ and $\mathscr{B}_m(i)$. Each link ij in the set of m mode-specific links, $ij \in \mathscr{L}_m$, is associated with a

choice. However, as we demonstrate in the next section, our estimates indicate that this assumption does not hold.

generalized link travel cost $t_{ij,m}$ which is specific to mode m and can be endogenously dependent on flow. Within this multimodal transport network, we define a primary network, $m = 1$, that represents the dense road network on which all cities and road intersections are located. All other modes, including rail and barges, are secondary transport networks ($m \neq 1$). A subset of nodes on the primary and secondary networks are intermodal terminals which allow for switches between mode-specific networks.

Consumption and Production A representative agent lives in location j, supplies one unit of labor inelastically, earns wage w_j, and consumes a continuum of varieties $\nu \in [0,1]$. Preferences are CES with elasticity of substitution $\sigma \geq 0$. We define aggregate income as Y^W, total labor endowment as $\bar{L}$, and normalize average income to one, so that $Y^W/\bar{L} = 1$.

Each location i produces a good $\nu \in [0,1]$ subject to a constant returns to scale technology and transports it to destination j along a feasible route. As we detail below, transport costs are route-specific in the sense that they depend on which neighboring node the good is being sourced from. Assuming perfect competition, the price of good ν in destination j from origin i along neighboring node k is given by $p_{ikj} = \frac{t_{kj}\tau_{ik}}{A_i} w_i$, where the marginal cost of production in i is $\frac{w_i}{A_i}$, local wages are w_i, each worker can produce A_i units of goods, and trade cost τ_{ik} is route-specific.[9]

Next, we endogenize τ_{ik} through routing, sourcing, and mode choice within the multimodal network.

3.2 Recursive Routing Choice

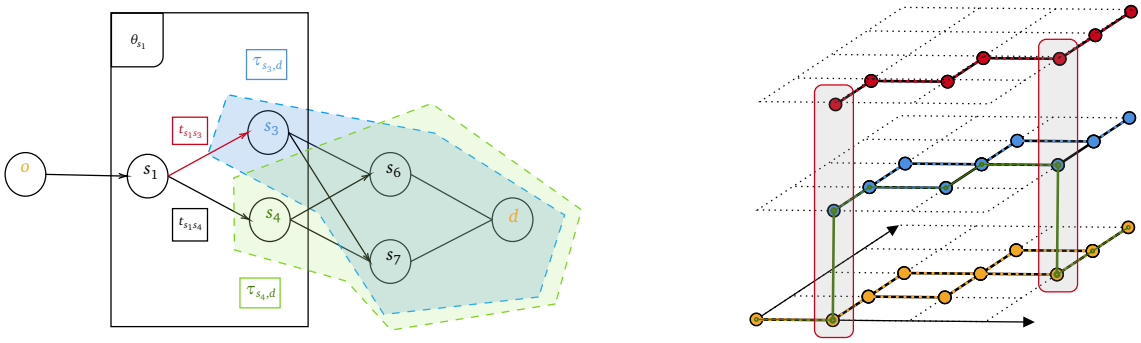
We begin with the routing block before extending the framework in the next subsection to encompass both routing and sourcing, building on recent advances in discrete route choice modeling (Daly and Bierlaire, 2006; Melo, 2012; Oyama, Hara and Akamatsu, 2022). Rather than enumerate all feasible paths, we represent route choice recursively.

The routing decision is a joint choice of elemental links in an ordered sequence from origin i to destination j (Papola and Marzano, 2013). At each node, the agent compares the cost of moving to each neighboring node—current link cost plus the continuation value from that node to the final destination. Figure 5a illustrates the idea. At node s_1, for example, the agent compares the cost of moving to s_3 and then continuing to d against the cost of moving to s_4 and then continuing to d.

More generally, at node i, the agent compares the cost of moving to any neighboring node $k \in \mathscr{F}(i)$, with link cost t_{ik}, and then continuing from k to destination j with continuation cost τ_{kj}. We assume

[9] In this paper, we assume perfect competition among freight transport companies. While this assumption simplifies our analysis, it also aligns with our focus on multimodal transport networks. Literature shows that multimodal container transport is generally more competitive than unimodal transport and that rail transport of containers is more competitive than non-containerized shipments. Zgonc, Tekavčič and Jakšič (2019) finds that multimodal road-rail transport remains competitive over short distances compared to unimodal road transport. Similarly, Surface Transportation Board (2009) reports lower markups for multimodal rail shipments and lower rail rates near alternative transport modes. Detailed analysis of market power in container freight transport would require comprehensive door-to-door shipment data, which is currently unavailable.

Figure 5. Recursive and Multimodal Routing



(a) Recursive Routing on a Directed Graph

(b) Multimodal Routing on a Multi-layered Graph

Notes: Panel (a) shows a directed transport network from origin o to destination d, with nodes as circles and edges representing transport costs. At node s_1, an agent chooses between neighboring nodes s_4 and s_3, each having a continuation value for minimizing transport cost to d. Panel (b) depicts multimodal routing on a multi-layered graph, similar to Panel (a), with nodes as circles and edges linked by mode-specific transport costs. Three mode-specific graphs are shown: primary roads (yellow), secondary rail and barges (blue and red). The overall figure shows a multi-layered graph with a green example route and intermodal transfers highlighted at nodes where mode choice occurs.

transport costs are multiplicative over links and that the agent experiences a segment-specific preference shock ε_{kj} that is iid Fréchet across successor nodes $k \in \mathscr{F}(i)$ with node-specific dispersion parameter θ_i.[10] Exploiting the properties of the Fréchet distribution, the expected transport cost from i to j is

$$\tau_{ij} = \mathrm{E}\left[\min_{k \in \mathscr{F}(i)} \left\{ \frac{t_{ik}\tau_{kj}}{\varepsilon_{kj}} \right\}\right] \propto \left(\sum_{k \in \mathscr{F}(i)} \left(t_{ik}\tau_{kj}\right)^{-\theta_i} \right)^{-\frac{1}{\theta_i}} \tag{3.1}$$

Equation (3.1) provides a recursive transport-cost index along the network topology.[11] It has three features worth emphasizing. First, transport costs are explicitly tied to network topology through the feasible-neighbor set $\mathscr{F}(i)$. In the Figure 5a example, the relevant choice set at s_1 consists only of the neighboring nodes s_3 and s_4. Second, allowing node-specific θ_i permits richer substitution patterns than the isoelastic benchmark. Third, with homogenous dispersion elasticities ($\theta_i = \theta$) and without considering multiple modes, this characterization is isomorphic to the routing characterization in AA2022,[12] while still retaining its analytically convenient feature.

[10] The recursive Fréchet formulation can equivalently be written as a Network Generalized Extreme Value (Network-GEV) model, where the observed topology of the network generates the relevant nesting structure. Appendix A.5 provides the mapping and the connection to Daly and Bierlaire (2006) and Lind and Ramondo (2023).

[11] The same routing problem can equivalently be written backward as a sourcing decision over predecessor nodes. Under homogeneous $\theta_i = \theta$, the two representations coincide; see Appendix A.5.

[12] This isomorphism between one-shot decisions over the universe of paths and sequential decision-making relying on dynamic programming has been previously established, most notably by Antràs and Gortari (2020) who explore a one-shot sequential sourcing decision in their main text and introduce a dynamic programming formulation in their Appendix A.1.3. Our work extends this formulation in two distinct ways: by explicitly formulating the decision-making on a graph, and by embedding the solution into the equilibrium conditions for a computationally convenient representation of the spatial equilibrium. See Appendix A.5 for further details.

3.3 Combined Routing and Sourcing Choice, and Recursive Equilibrium

We now combine the recursive routing choice in Equation (3.1) with the stochastic sourcing choice from Eaton and Kortum (2002) to derive a closed-form characterization for both route and sourcing choices. Next, we integrate the route and sourcing choice into the equilibrium conditions, relating endogenous equilibrium outcomes recursively along the network topology.

Routing and Sourcing Under perfect competition, consumers (recursively) source their products by choosing the lowest-cost route-source combination. The price the consumer in j faces is the expected minimum of choosing a variety ν that is both the least cost source i and being routed along the lowest cost route from i to j utilizing the neighboring node k,

$$p_j \equiv \mathbf{E}_\varepsilon \left[\min_{(k,i)\in\mathscr{B}(j)\times N} \frac{t_{kj}\tau_{ik}}{\varepsilon_{ikj}(\nu)} w_i \right] \tag{3.2}$$

where, following Eaton and Kortum (2002), we assume that $\varepsilon_{ikj}(\nu)$ is iid Fréchet distributed across the set of (predecessor) neighboring nodes $k \in \mathscr{B}(j)$ and sourcing partners i with scale parameter $1/A_i$, where A_i captures origin-specific efficiency and θ_j is a node-specific shape parameter which regulates the inverse of shock dispersion.[13] Given the distribution of shocks, the probability that j sources a good from i that is routed via neighboring node k is as follows (see detailed derivations in Appendix A.1),

$$\underbrace{\pi_{ij,k}}_{\substack{\text{Joint routing-sourcing prob.}\\ \text{from } i \text{ to } j \text{ via node } k}} = \underbrace{\frac{\tau_{ij}^{-\theta_j}\left(\frac{w_i}{A_i}\right)^{-\theta_j}}{\sum_{i'\in N}\tau_{i'j}^{-\theta_j}\left(\frac{w_{i'}}{A_{i'}}\right)^{-\theta_j}}}_{\text{Sourcing prob. from } i} \times \underbrace{\frac{\left(t_{ik}\tau_{kj}\right)^{-\theta_j}}{\sum_{k'\in\mathscr{B}(j)}\left(t_{ik'}\tau_{k'j}\right)^{-\theta_j}}}_{\text{Routing prob. via node } k} \equiv \pi_{ij} \times \pi_{ij}^{ik} \tag{3.3}$$

where we have decomposed the combined routing-sourcing probability into two different components. The first term, π_{ij}, represents the sourcing probability given the expected bilateral transport cost, τ_{ij}, defined as and is identical to the sourcing probability in Eaton and Kortum (2002) except for the inclusion of heterogeneous substitution elasticities along the network θ_j. The second term, π_{ij}^{ik}, represents an implicit route choice probability, characterizing the likelihood of choosing a neighboring node k amongst all neighboring nodes when sourcing from origin i.

Recursive Equilibrium Using Equation (3.3), we derive a generalized gravity equation and then impose goods-market clearing and balanced trade (see Appendix A.2 and Appendix Section A.3). The resulting equilibrium conditions are:

[13] The Frechet distribution's properties imply that the inclusive value is given by $p_j = \mathbb{E}_\varepsilon\left[\min_{(k,i)\in\mathscr{B}(j)\times N}\frac{t_{kj}\tau_{ik}w_i}{\varepsilon_{ij,k}(\nu)}\right] = \left(\sum_{(k,i)\in\mathscr{B}(j)\times N}\left(t_{kj}\tau_{ik}w_i\right)^{-\theta_j}\right)^{-\frac{1}{\theta_j}} = \left(\sum_{k\in\mathscr{B}(j)}t_{kj}^{-\theta_j}\sum_{i\in N}\left(\tau_{ik}w_i\right)^{-\theta_j}\right)^{-\frac{1}{\theta_j}}$. Note that in the last equation, we can re-arrange the sums to highlight that under a homogenous dispersion parameter ($\theta_j = \theta$), the formula takes on a recursive form and connects expected sourcing prices back to origin i along the network topology, $p_j^{-\theta} = \sum_{k\in\mathscr{N}(j)} t_{kj}^{-\theta} p_k^{-\theta}$ since $p_k^{-\theta} = \sum_{i\in N}\left(\tau_{ik}w_i\right)^{-\theta}$.

$$\Pi_i^{-\theta_i} = \left(t_{ii}^{-\theta_i}\right)\frac{\delta_i}{P_i^{-\theta_i}} + \sum_{k\in\mathcal{F}(i)} t_{ik}^{-\theta_i}\tilde{\Pi}_k^{-\theta_i}, \qquad P_j^{-\theta_j} = \left(t_{jj}^{-\theta_j}\right)\frac{\gamma_j}{\Pi_j^{-\theta_j}} + \sum_{k\in\mathcal{B}(j)} t_{kj}^{-\theta_j}P_k^{-\theta_j}, \tag{3.4}$$

where $\{\Pi_i, P_i\}$ are the producer and consumer market-access terms.[14]

We can reformulate the equilibrium equations to solve for the spatial allocation of labor and income, allowing for a rich set of amenity and productivity spillovers. First, we assume welfare equalization, given by $W_j = \frac{w_j}{P_j}u_j$. Then, we allow for localized productivity (A_i) and amenity spillovers (u_i) that depend on the density of workers in a locality. Specifically, productivity is modeled as $A_i = \bar{A}_i L_i^{\alpha}$ and amenities as $u_i = \bar{u}_i L_i^{\beta}$, where $\bar{A}_i$ represents the exogenous productivity component at location i, and α measures the sensitivity of productivity to local population L_i (capturing productivity spillovers). Similarly, $\bar{u}_i$ denotes the exogenous utility from residing in location i, while β indicates the extent to which amenities are influenced by the local population (capturing amenity spillovers). This allows us to rewrite the price indices in terms of labor and income allocations (Footnote (1) in Appendix A.2). As a result, we have consumer price index $P_j = \frac{1}{\bar{W}}\bar{u}_j L_j^{\beta-1}Y_j$ and producer price index $\Pi_i = \bar{A}_i L_i^{1+\alpha}Y_i^{-\frac{\theta_i+1}{\theta_i}}$.[15] Substituting both price indices into the equilibrium system in Equation (3.4), we obtain a system of equation that determines the the spatial distribution of income and labor as well as a global scalar that determines aggregate welfare $\bar{W}$ (see Appendix Section A.3 for further details).

Specifically, given the productivity and amenity fundamentals $\left\{\bar{A}_i, \bar{u}_i\right\}$ as well as the aggregate transport costs, the system of $2N$ equations can be solved for the $2N$ endogenous equilibrium values. The equilibrium system determines the endogenous variables via the interaction of price indices along the network topology, where transport cost is endogenously determined as part of the equilibrium system.

Our framework has three main benefits. First, our approach allows us to nest the choice of different transport modes within the optimal route choice model, with a separate elasticity of substitution across modes from the elasticity across routes. This is possible because our model recasts the equilibrium problem in terms of edge-level outcomes and allows us to incorporate richer edge-level decision making and substitution margins. Second, we can include a rich set of mode-specific congestion patterns, including congestion at intermodal terminals as well as along the primary network. Third, we do not rely on matrix inversion methods and can derive closed-form expressions for equilibrium transport costs, traffic flows, and market access terms, which depends on mode and route choices as well as the topology of the transport network. To focus on multimodal transport networks and incorporate transport mode choice into this framework, we assume a homogeneous route choice elasticity ($\theta_j = \theta$) from now.

[14] Here $\tilde{\Pi}_k$ is an auxiliary producer-access object that adjusts for heterogeneity in θ_j. It collapses to Π_k under homogeneous $\theta_j = \theta$. See Appendix A.3 for the exact expression and derivation.

[15] We can furthermore rewrite the equations in terms of rescaled variables, $\{y_i, l_i\}$, where we define shares of world income in location i, $y_i \equiv \frac{Y_i}{Y^W}$, and shares of total labor in location i, $l_i \equiv \frac{L_i}{L^W}$. To illustrate this, we did this explicitly for the counterfactual equilibrium system in the Appendix A.10.

3.4 Nested Mode Choice and Aggregate Transport Cost

Previously, we considered a consumer making a route-source decision recursively, where choices are made sequentially along links of the graph. To expand this framework to incorporate routing decisions across multiple transport modes, we introduce a nested mode-choice structure into the current setup. Now, given the node the consumer has selected, she then chooses the cost-minimizing mode out of all available mode options when travelling to that node (subject to an extreme value distributed cost shock). To fix ideas, consider a consumer in location i who has chosen to route via neighboring node k. At k, she then considers the different modes that are available along link ik, i.e. $m \in \mathscr{M}_{ik}$, where $\mathscr{M}_{ik}$ is the set of modes available between i and k. One way to represent this is using a multi-layered graph (Figure 5b), where the set of available nodes depends on the existing infrastructure. Each layer in Figure 5b represent a mode-specific network, with the top red layer having fewer nodes than the blue (middle) and yellow (bottom) layers. The consumer selects a mode-specific layer for each link in the network.

In this setting we assume that the link-level and mode-specific transport costs for the primary mode is $t_{ik,1}$, but for any secondary mode ($m \neq 1$) a switching cost is imposed, i.e. $t_{ik,m} = s_{ii,m}\iota_{ik,m}s_{kk,m}$ for $m \neq 1$, where $\iota_{ik,m}$ refers to the iceberg transport cost of traversing the link between node i and k along mode m. Notice that this specification is general and allows for geographies where the secondary modes might connect both primary and secondary nodes. We present the expected aggregate transport cost as the sum of the endogenous mode-specific transport cost, $t_{ik,m}$, that is available between neighboring nodes i and k, i.e. $\mathscr{M}(i,k)$ (see detailed derivations in Appendix A.4):

$$t_{ik} = \mathbb{E}\left[\min_{m\in\mathscr{M}(i,k)}\left\{\frac{t_{ik,m}}{\varepsilon_{ik,m}}\right\}\right] \propto \left(\sum_{m\in\mathscr{M}(i,k)} t_{ik,m}^{-\eta}\right)^{-\frac{1}{\eta}} \tag{3.5}$$

where the mode choice is subject to an extreme-value distributed shock with dispersion parameter η. The corresponding mode share is:

$$\pi_{ik}^{m} = \frac{t_{ik,m}^{-\eta}}{t_{ik}^{-\eta}} \tag{3.6}$$

This nested structure introduces a second substitution margin without altering the recursive structure of the equilibrium system. The key distinction is that substitution across modes is governed by η, while substitution across routes and sources is governed by θ. When $\eta = \theta$, or when only one transport mode is available, the model collapses to the single-elasticity case nested by AA2022 (Appendix A.5).

This setup allows for a flexible introduction of transport mode choice in a setting where agents are already making endogenous routing and sourcing decisions. A lower cost for one mode affects behavior in two ways. Directly, it makes that mode relatively more attractive on the affected link. Indirectly, because the aggregate cost of the link falls, it changes the attractiveness of routes and locations that use that link. This distinction between within-link modal substitution and broader route-level reorganization will be important below.

3.5 Congestion and Traffic Flows

To introduce congestion, we proceed in three steps. First, we characterize aggregate traffic on the network. Second, we derive mode-specific and terminal traffic. Third, we let these traffic objects feed back into transport costs through road and terminal congestion.

Traffic Flows on the Aggregate Network We start by constructing a measure of link intensity—the probability that route i to j is used and the likelihood of selecting a specific link kl on that route:

$$\pi_{ij}^{kl} = \left(\frac{\tau_{ij}}{\tau_{ik} t_{kl} \tau_{ij}} \right)^{\theta} \tag{3.7}$$

This formula follows directly from our recursive routing choice framework (see Appendix A.6.1). We then use these derivations to define a gravity equation for aggregate traffic flows on link kl,

$$\Xi_{kl} = t_{kl}^{-\theta} \times P_k^{-\theta} \times \Pi_l^{-\theta} \tag{3.8}$$

where traffic flows are a function of inward, $P_k^{-\theta}$, and outward market access measures, $\Pi_l^{-\theta}$, as well as the aggregate transport cost for that link across the multimodal transport network.

Traffic Flows on the Mode-Specific Networks The nested choice implies that the probability of using route ij, selecting link kl, and choosing mode m on that link is

$$\pi_{ij,kl,m} = \pi_{ij} \times \pi_{ij}^{kl} \times \pi_{kl}^{m},$$

where π_{ij} is the joint source-route share from Equation (3.3), π_{ij}^{kl} is the link-use probability from Equation (3.7), and $\pi_{kl}^{m} = t_{kl,m}^{-\eta} / t_{kl}^{-\eta}$ is the conditional mode share on link kl. Summing over origins and destinations yields mode-specific traffic (see Appendix Section A.6.2 for details),

$$\Xi_{kl,m} = \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} \pi_{ij,kl,m} E_j = t_{kl,m}^{-\eta} \times t_{kl}^{\eta-\theta} \times P_k^{-\theta} \times \Pi_l^{-\theta} \tag{3.9}$$

which shows that the mode-specific traffic is the conditional mode share of aggregate traffic, i.e. $\Xi_{kl,m} = \pi_{kl}^{m} \Xi_{kl}$. This expression makes the two layers of the model transparent. First, aggregate traffic depends on the composite link cost and market access. Mode-specific traffic then allocates this aggregate flow across modes according to their relative costs.

Traffic Flows at Terminals We next characterize traffic at terminals where goods switch between the primary and secondary modal networks. Summing mode-specific traffic over all successor nodes reachable through terminal k on mode m yields (see Appendix A.6.3 for details),

$$\Xi_{kk,m} = P_k^{-\theta} \times s_{kk,m}^{-\eta} \times \sum_{l \in \mathcal{F}_m(k)} \left(\tau_{kl,m} s_{ll,m} \right)^{-\eta} \times t_{kl}^{\eta-\theta} \times \Pi_l^{-\theta}, \tag{3.10}$$

This equation differs slightly from the two previous traffic equations due to an additional summation term. This term, a higher-order market access measure, shows how traffic flows through a terminal depends on the sum of traffic generated by nodes that can be accessed via that terminal along the

secondary network. This term also captures the centrality of terminals in connecting primary and secondary networks, highlighting their potential to become bottlenecks in the overall transport network.

Congestion on the Primary Transport Network. Having derived traffic flows, we now incorporate congestion on the primary network and then at terminal stations.[16] The congestion model for the primary network is based on AA2022, assuming that travel costs on a link depend on its traffic volume,

$$t_{kl} = \bar{t}_{kl}\left(\Xi_{kl,1}\right)^{\lambda_1}, \tag{3.11}$$

where λ_1 determines the strength of congestion on the primary network, $\overline{\mathbf{T}} \equiv [\bar{t}_{kl}]$ is the primary network's infrastructure, and $\Xi_{kl,1}$ is the network's link-level traffic flows. As long as $\lambda_1 > 0$, transport cost on a link increases as its traffic flow increases.[17] Using this expression, we first state the primary network's transport costs as a function of its infrastructure, aggregate-level transport costs, and market access terms ($t_{kl,1}$, first equation below). Next, we combine this transport cost with Equation (3.8) to express the equilibrium traffic flows ($\Xi_{kl,1}$, second equation below):

$$t_{kl,1} = \bar{t}_{kl,1}^{\frac{1}{1+\eta\lambda_1}} \times t_{kl}^{\frac{\lambda_1(\eta-\theta)}{1+\eta\lambda_1}} \times P_k^{\frac{-\theta\lambda_1}{1+\eta\lambda_1}} \times \Pi_l^{\frac{-\theta\lambda_1}{1+\eta\lambda_1}}, \quad \Xi_{kl,1} = \bar{t}_{kl,1}^{\frac{-\eta}{1+\eta\lambda_1}} \times t_{kl}^{\frac{\eta-\theta}{1+\eta\lambda_1}} \times P_k^{\frac{-\theta}{1+\eta\lambda_1}} \times \Pi_l^{\frac{-\theta}{1+\eta\lambda_1}} \tag{3.12}$$

The equilibrium traffic flows on the primary network depend on the fundamental transport capacity and aggregate transport cost of each link ($\bar{t}_{kl,1}$ and t_{kl}), both inward and outward market access terms, as well as the strength of the congestion externality (λ_1). Intuitively, as better market access improves traffic flow on each link, this also increases congestion which mutes its overall impact.

Congestion at Intermodal Terminals. When introducing congestion at terminals, we assume that the direct cost of transiting through a terminal depends on the overall traffic at the terminal,

$$s_{kk,m} = \bar{s}_{kk,m}\left[\Xi_{kk,m}\right]^{\lambda_m}, \quad \text{where} \quad m \neq 1 \tag{3.13}$$

where λ_m specifies the strength of congestion at terminals, $\overline{\mathbf{S}} \equiv \left[\bar{s}_{kk,m}\right]$ is the switching matrix that connects primary and secondary modes, and $\Xi_{kk,m}$ denotes the traffic at the terminal station. Combining with the terminal traffic expression (Equation (3.10)), the transport cost for secondary modes, $t_{kl,m}$, is

$$t_{kl,m} = \bar{s}_{kk,m} \times \tau_{kl,m} \times \bar{s}_{ll,m} \times P_k^{-\theta\lambda_m} \times \Pi_l^{-\theta\lambda_m} \times P_{k,m}^{-\theta\lambda_m} \times \Pi_{l,m}^{-\theta\lambda_m}, \quad \text{where} \quad m \neq 1 \tag{3.14}$$

where $P_{l,m} \equiv \sum_{k\in\mathscr{B}(l)} t_{kl,m}^{-\eta} \times t_{kl}^{\eta-\theta} \times P_k^{-\theta}$ and $\Pi_{k,m} \equiv \sum_{l\in\mathscr{F}(k)} t_{kl,m}^{-\eta} \times t_{kl}^{\eta-\theta} \times \Pi_l^{-\theta}$ define mode-specific market access measures. Intuitively, if $\lambda_m > 0$, transport cost increases with increasing traffic at the terminal. By combining Equation (3.14) with our modal traffic expression (Equation (3.9)), we determine

[16] While it is theoretically feasible to extend the framework to include congestion on the secondary network, we have opted to model congestion at terminals. This choice reflects the fact that bottlenecks then to occur at the intermodal transfer with a prime example being port congestion. Additionally, incorporating congestion at the route-level for secondary networks would come at the cost of added complexity and decreased traceability.

[17] While somewhat different to the more commonly used Bureau of Public Roads (BPR) function (Boyles, Lownes and Unnikrishnan, 2021), it is both analytically convenient and can be micro-founded in a simple model where transport costs are log-linear in travel time and speed is a log-linear function of traffic congestion as shown in AA2022.

the traffic flows for the secondary modes while taking into account endogenous congestion at terminals,

$$\Xi_{kl,m} = \bar{s}_{kk,m}^{-\eta} \times \tau_{kl,m}^{-\eta} \times \bar{s}_{ll,m}^{-\eta} \times P_{k,m}^{\eta\theta\lambda_m} \times \Pi_{l,m}^{\eta\theta\lambda_m} \times t_{kl}^{\eta-\theta} \times P_k^{-\theta(1-\eta\lambda_m)} \times \Pi_l^{-\theta(1-\eta\lambda_m)}, \quad \text{where} \quad m \neq 1 \tag{3.15}$$

While increases in both incoming and outgoing market access measures generally increases traffic flows (Equation (3.8)), we highlight how these market access terms contribute to traffic flows on both the primary and secondary networks while taking congestion into account (Equations (3.12) and (3.15) respectively). Specifically, terminal congestion (λ_m) can reduce the appeal of multimodal routes and consequently decrease traffic on the secondary network (Equation (3.15)). This highlights the significant role terminals can have as potential bottlenecks within a multimodal network. In Section 4.2, we estimate the terminal congestion elasticity by focusing on one important type of intermodal terminal: ports.

3.6 Decomposing the Qualitative Channels

To build intuition, this section characterizes how transport cost changes propagate through the network, affecting traffic and welfare. A lower mode-specific transport cost on link kl first changes traffic allocation across modes on that link and then across routes in the network. These adjustments then determine whether the resulting traffic increase is absorbed by congestion across links and at terminals, or converted into welfare gains through better connectivity and market access. The recursive equilibrium system developed earlier in Section 3.2 allows for a particularly convenient application of the inverse function theorem to characterize welfare elasticities in closed-form, following Allen, Fuchs and Wong (2025). We extend these results further to derive decompositions of the welfare elasticities, with the goal of isolating the contributions of the four main ingredients in our model: road and terminal congestion, productivity and amenity spillovers, multiple modes, and endogenous routing.

Traffic Adjustments: Substitution and Complementarity Across Modes We start with the traffic adjustments. On link kl, mode-specific traffic $\Xi_{kl,m}$can be written as

$$\Xi_{kl,m} = \pi_{kl}^m \Xi_{kl}, \qquad \pi_{kl}^m \equiv \frac{t_{kl,m}^{-\eta}}{t_{kl}^{-\eta}}, \qquad \Xi_{kl} \equiv t_{kl}^{-\theta} P_k^{-\theta} \Pi_l^{-\theta}, \tag{3.16}$$

where $t_{kl,m}$ is the cost of moving on link kl using mode m, t_{kl} is the corresponding aggregate link cost, η is the elasticity of modal substitution, θ is the elasticity governing route and trade substitution, while P_k and Π_l are outgoing and incoming market access terms at k and l. A mode-specific cost change therefore works through two traffic-side margins. First, it changes the conditional mode share on link kl (π_{kl}^m), so traffic is reallocated across modes on that link. Second, it changes the aggregate cost of using that link, which changes the market access of the locations connected to it and the total traffic on that link (Ξ_{kl}). Appendix A.7 shows that, for an alternative mode $m \neq m'$, mode-specific transport cost changes can

generate the following direct and indirect traffic adjustments:

$$\frac{d \ln \Xi_{kl,m}}{d \ln t_{kl,m'}} \approx \Big[\underbrace{\eta}_{\text{direct substitution}} \underbrace{- \theta\big(1 + [\widetilde{\Omega}^{t,P}]_{kl} + [\widetilde{\Omega}^{t,\Pi}]_{kl}\big)}_{\text{indirect complementarity}} \Big] \varpi_{kl,m'}, \tag{3.17}$$

where $\varpi_{kl,m'} \equiv \Xi_{kl,m'} / \sum_n \Xi_{kl,n}$ is the baseline traffic share of mode m' on link kl, and $\widetilde{\Omega}^{t,P}$ and $\widetilde{\Omega}^{t,\Pi}$ are reduced-form centrality objects that summarize how a change in the cost of link kl propagates into inward and outward market access.

The first term in Equation (3.17) captures direct *modal substitution*: if mode m' becomes cheaper, traffic shifts toward it and away from alternative mode m on the same link. In the example from the introduction, improving a link on the road network reduces road transport costs, making rail transport relatively more expensive and shifting traffic from rail to road on that link. The second term captures the broader indirect *modal complementarity* effect: because the aggregate cost of link kl falls, that link becomes more attractive within the network as a whole, which raises total traffic and market access and can increase traffic on other modes as well. Better road access can thereby increase transport demand across all modes including rail. The strength of this latter channel is stronger for links that are more central within the network. We now turn to the welfare implications of these traffic adjustments.

Welfare Decomposition: Congestion and Reallocation Wedges The traffic expression above in Equation (3.17) studies a *mode-specific* perturbation $d \ln t_{kl,m'}$. Here, the welfare decomposition values that same kind of local transport change through the welfare elasticity of a marginal reduction in the primitive mode-specific cost $\bar{t}_{kl,m}$ on link kl. Under the common-baseline convention used throughout, the mode-specific shock is first mapped into the composite cost of link kl and then propagated through the same equilibrium system under different model variants—each shutting down one main ingredient in our model in order to isolate its contribution. Appendix A.8 shows that this mapping is proportional to the conditional mode share on the link, so the normalized multiplier remains link-specific while the exposure term remains mode-specific. With that interpretation in mind, let

$$E^c_{kl,m} \equiv -\frac{d \ln W^c}{d \ln \bar{t}_{kl,m}} = \rho \, \Xi_{kl,m} \, m^c_{kl}, \tag{3.18}$$

where $E^c_{kl,m}$ is the local welfare elasticity under model variant c, $\Xi_{kl,m}$ is the observed baseline traffic of mode m on link kl, ρ is a common scaling term, and m^c_{kl} is the exact local multiplier that summarizes how the same initial mode-specific link improvement works its way through the network under that variant.[18] $\Xi_{kl,m}$ tells us how exposed the economy is to the improved mode-specific link, while m^c_{kl} tells us whether the resulting extra traffic is reallocated toward more valuable routes and modes, runs into congestion across links or at terminals, or generates broader market-access gains.

[18] Appendix A.8 writes this multiplier as $m^c_{kl} = \omega^\top J_c^{-1} s_{kl,m}$, where J_c is the Jacobian of the transformed equilibrium system under model variant c, $s_{kl,m}$ is the normalized shock created by lowering the primitive cost of mode m on link kl, and ω maps the induced equilibrium responses into welfare. Under the reduced equilibrium representation, $s_{kl,m} = s_{kl}$, so the normalized multiplier depends on the link rather than on the mode.

To individually isolate the contributions of four main mechanisms in our model, we compare the full model F to the following model variants ($c \in \{\text{NC, NT, RO, FR, H}\}$): a no-congestion benchmark NC, a no-terminal-congestion benchmark NT, a road-only benchmark RO that suppresses multimodal flexibility, a fixed-routing benchmark FR that suppresses endogenous rerouting, and a Hulten benchmark that shuts down agglomeration and dispersion externalities H.[19] Across all of these comparisons, the same initial mode-specific link improvement is evaluated at the observed transport assignment; what changes is not the shock itself, but which margins are allowed to respond after it hits. We define the four transport-side wedges as follows,

$$\underbrace{d_{kl}^{\text{road}} \equiv m_{kl}^{\text{NC}} - m_{kl}^{\text{NT}}, \quad d_{kl}^{\text{term}} \equiv m_{kl}^{\text{NT}} - m_{kl}^{F}}_{\text{Congestion Wedges}}, \qquad \underbrace{d_{kl}^{\text{mode}} \equiv m_{kl}^{F} - m_{kl}^{\text{RO}}, \quad d_{kl}^{\text{route}} \equiv m_{kl}^{F} - m_{kl}^{\text{FR}}}_{\text{Reallocation Wedges}} \tag{3.19}$$

where d_{kl}^{road} and d_{kl}^{term} are the *congestion wedges*: they measure how much of the potential gains are absorbed once congestion, across links and at terminals, is allowed to bind. Specifically, road congestion wedge d_{kl}^{road} is defined as the difference between no-congestion benchmark m_{kl}^{NC} and no-terminal-congestion benchmark m_{kl}^{NT}, isolating the welfare loss due to link-level road congestion. The terminal congestion wedge d_{kl}^{term} is the difference between m_{kl}^{NT} and the full model m_{kl}^{F}, capturing terminal bottlenecks. In contrast, d_{kl}^{mode} and d_{kl}^{route} are the *reallocation wedges*: they measure whether the way traffic reallocates across modes and routes makes the improved link more valuable or easier to substitute away from. The mode reallocation wedge d_{kl}^{mode} is defined as the difference between the full model m_{kl}^{F}—with multiple modes—and the road-only benchmark m_{kl}^{RO} while the route reallocation wedge d_{kl}^{route} is defined as the difference between the full model m_{kl}^{F} and the fixed-routing benchmark m_{kl}^{FR}.

The welfare effects of the full model can be decomposed into distinct qualitative channels forming an explicit decompositional "ladder" over the set of benchmark models:

$$\underbrace{m_{kl}^{F}}_{\substack{\text{Full}\\ \text{Model}}} = \underbrace{m_{kl}^{\text{NC}}}_{\substack{\text{No-Congestion}\\ \text{Model}}} - d_{kl}^{\text{road}} - d_{kl}^{\text{term}} = \underbrace{m_{kl}^{\text{NT}}}_{\substack{\text{No-Terminal}\\ \text{Congestion Model}}} - d_{kl}^{\text{term}} = \underbrace{m_{kl}^{\text{RO}}}_{\substack{\text{Road-Only}\\ \text{Model}}} + d_{kl}^{\text{mode}} = \underbrace{m_{kl}^{\text{FR}}}_{\substack{\text{Fixed-}\\ \text{Routing Model}}} + d_{kl}^{\text{route}}. \tag{3.20}$$

In addition to this transport-side ladder, it is useful to keep the efficient-equilibrium Hulten (1978) benchmark in view.[20] This is the natural comparison point for our Hulten terms: it is the benchmark

[19] An interesting comparison point is Donald, Fukui and Miyauchi (forthcoming). They ask a broad welfare-accounting question: in a utilitarian spatial model with heterogeneous households, idiosyncratic location preferences, and transfers, what primitive wedges make welfare depart from the Fogel–Hulten benchmark? Their answer is a decomposition into technology, marginal-utility dispersion, fiscal externalities, technological externalities, and redistribution. Our focus is different. We specialize to the utility-equalization transport benchmark and use the recursive multimodal network structure to decompose the model-implied welfare elasticity itself. The benefit of our approach is that it turns the extended social-savings formula from a black box into a mechanism-level anatomy: we can say whether a link's welfare value is being attenuated by road or terminal scarcity, amplified or offset by modal diversion and rerouting, or propagated through broader network feedback. In that sense, Donald, Fukui and Miyauchi (forthcoming) classify deviations from Hulten by welfare primitive, whereas we classify them by transport mechanism. We consider the two perspectives as complementary.

[20] Allen, Fuchs and Wong (2025) show that, when congestion and agglomeration/dispersion externalities are shut down so that the equilibrium is efficient, a Hulten-type envelope theorem applies and the welfare elasticity of a marginal link-mode improvement is proportional to the observed traffic share on that link-mode. In the notation of this paper, that efficient-equilibrium benchmark is $E_{kl,m}^{H} = \Xi_{kl,m}$, or equivalently $m_{kl}^{H} = 1/\rho$.

valuing the welfare effect of a local transport improvement solely by its baseline traffic while excluding agglomeration and dispersion externalities (productivity and amenity effects). The benchmark gap is

$$h_{kl} \equiv \frac{E^H_{kl,m} - E^F_{kl,m}}{\Xi_{kl,m}} = \rho\left(m^H_{kl} - m^F_{kl}\right) = (1-\rho) + \rho(1 - m^F_{kl}), \qquad \Delta^{\text{Hulten}}_{kl,m} = \Xi_{kl,m} h_{kl}. \tag{3.21}$$

Unlike d^{road}_{kl}, d^{term}_{kl}, d^{mode}_{kl}, and d^{route}_{kl}, the Hulten gap is not a comparison between two transport-side Jacobian variants. Instead, it benchmarks the full equilibrium multiplier against the efficient-equilibrium Hulten object in which the welfare gain is proportional to the baseline traffic share. Economically, h_{kl} measures the departure of the full-model welfare elasticity from that efficient benchmark: it is positive when congestion and externality feedback attenuate the gain relative to the traffic-share benchmark, and negative when network and spillover feedback amplify the gain beyond it.

We discuss each of these wedges qualitatively (formal characterization in Appendix A.8). Proposition 5 shows that the road and terminal *congestion wedges* can be written as the interaction between two things: (1) how much a mode-specific link improvement increases traffic on congested road or terminals, and (2) how valuable it is to relieve congestion on those margins. Economically, they capture congestion losses: when improvements push traffic toward already congested road links or heavily used terminals, part of the traffic-side complementarity in Equation (3.17) is absorbed by congestion elsewhere in the network rather than passed through to welfare.

Turning to the mode and route *reallocation wedges*, Proposition 6 shows that each wedge can be decomposed into three parts:

$$d^{\text{mode}}_{kl} = d^{\text{mode}}_{kl,\text{alloc}} + d^{\text{mode}}_{kl,\text{cong}} + d^{\text{mode}}_{kl,\text{glob}}, \qquad d^{\text{route}}_{kl} = d^{\text{route}}_{kl,\text{alloc}} + d^{\text{route}}_{kl,\text{cong}} + d^{\text{route}}_{kl,\text{glob}}. \tag{3.22}$$

where the first term captures local reallocation effects: how lower costs on link kl change nearby mode use and routing ($d^{\text{mode}}_{kl,\text{alloc}}$ and $d^{\text{route}}_{kl,\text{alloc}}$). The second term captures the resulting congestion feedback from the reallocation ($d^{\text{mode}}_{kl,\text{cong}}$ and $d^{\text{route}}_{kl,\text{cong}}$), while the third term captures broader spatial feedback as changes in transport access reallocate activity and demand across locations through market access channels.

This three-way decomposition sharpens the economic interpretation of the *reallocation wedges*. An overall positive mode wedge means that multimodal flexibility raises the value of the mode-specific link improvement because modal reallocation and its congestion consequences reinforce the improvement, and broader equilibrium feedback does not offset that effect. A negative mode wedge means that extra modal flexibility mainly allows traffic to substitute away from the improved link. From the earlier example, a positive mode wedge means that multimodal flexibility raises the value of a road-link improvement by expanding overall transport activity on that link, including rail, while a negative mode wedge means that this flexibility reduces reliance on the improved link, lowering its marginal welfare value.

For the route reallocation wedge, a positive wedge means that endogenous rerouting reinforces the value of the improved link by shifting traffic onto complementary paths that make the improved link more central within the network. A negative route wedge instead implies that rerouting reduces the

network's dependence on the improved link by opening up alternative paths. For example, a road-link improvement can have a positive route wedge if rerouting directs traffic toward other road or rail links that are connected to the improved road, but a negative route wedge if it diverts traffic toward substitute corridors that diminish the role of that road segment.

We further provide conditions under which the wedges can be signed. Under the one-sided conditions reported in Appendix A.8.4, the *congestion wedges* are positive because additional traffic on already congested links or at terminals lowers the pass-through from a transport improvement to welfare. For the *reallocation wedges*, the appendix provides sufficient conditions and a dominance criterion: they are positive when the value created by local route and mode reallocation, together with the associated congestion relief, is larger than the offsetting aggregate feedback term.

Taken together, the section says one thing. A cheaper link first reorganizes traffic. That reorganization then either runs into congestion across links or at terminals, or increases the usefulness of the link within the broader network. The traffic decomposition identifies where the shock creates substitution and complementarity, while the welfare decomposition shows whether those same adjustments appear mainly as *congestion wedges* or as *reallocation wedges*.

4 Estimation: Modal Substitution and Terminal Congestion

In order to quantify our model, we require two key elasticities. First, to motivate the mode choice channels in our model, we require an elasticity of substitution between transport modes (η). In Section 4.1 we estimate this elasticity by examining how traffic shifts between modes in response to cost changes in a one mode. Second, the counterfactual equilibrium in our model depends on the strength of the congestion forces within the network, particularly at intermodal terminals (λ_m).[21] In Section 4.2, we estimate λ_m by analyzing the relationship between ship dwell times and port traffic. We demonstrate that this elasticity, though estimated using port terminals, also applies to inland intermodal rail terminals.

4.1 Estimating the Elasticity of Modal Substitution

We start by motivating our approach to estimate the elasticity of modal substitution η, and its structural interpretation through the lens of our model, as discussed in Section 3.6. Next, we explain how we estimate this elasticity by adopting the research design from Duranton and Turner (2011) and incorporating the focus on relative modal responses from Hummels and Schaur (2013) and Lugovskyy, Skiba and Terner (2022). Third, we conduct several robustness checks to validate our results.

First, consider a mode-specific transport cost decrease, such as a local road improvement. As discussed

[21] There are two congestion forces in our model: on the primary network and at intermodal terminals (λ_1, λ_m). In this paper, we rely on prior literature for λ_1 and focus on estimating λ_m in this paper.

in Subsection 3.6, this improvement has both direct and indirect effects on traffic. Directly, lower road costs will substitute traffic from rail or barges to road. The strength of this substitution effect directly depends on the parameter of interest η. Indirectly, lower road costs will reduce aggregate trade costs and improve overall market access across all modes (the complementarity effect). Since we are interested in estimating η, the complementarity effect constitutes confounding variation, as changes in mode-specific traffic are endogenous to changes in both incoming and outgoing market access (Equation (3.17)). We address this, and demonstrate more formally in the Appendix B.2, by examining the instrumented changes in relative modal traffic responses (modal split) to mode-specific cost changes, rather than through the mode-specific traffic responses directly (Equation (B.3), Appendix B.2).

Empirical Strategy Our approach builds on Duranton and Turner (2011) (hereafter abbreviated to DT2011), who found that improving road infrastructure by increasing lane kilometers of interstate highways, increases the vehicle-kilometers traveled (VKT) in U.S. metropolitan cities, particularly for commercial freight trucks. To address the endogeneity between VKT demand and road stock changes at the metropolitan statistical area (MSA) level, the paper utilizes three instruments: kilometers of preliminary interstate highway from the 1947 highway plan in each MSA (Baum-Snow, 2007; Michaels, 2008), kilometers of 1898 railroads in each MSA, and 1528-1850 exploration routes.[22]

By aggregating our waybill rail traffic data to the MSA level and matching it to the DT2011 data, we are able to estimate the elasticity of modal substitution. We do this by studying how traffic shifts between modes in response to cost changes in one mode, specifically the effect of reduced road costs via interstate highway improvements on relative rail and truck road traffic flows.[23] We start with the following OLS specification with fixed effects from DT2011:

$$\ln Y_{cy} = \alpha \ln \text{Interstate Highway Lanes}_{cy} + \psi C_{cy} + \zeta_c + \iota_y + \varepsilon_{cy} \tag{4.1}$$

where $\ln Y_{cy}$ is log traffic use outcomes for MSA c in year y, and $\ln \text{Interstate Highway Lanes}_{cy}$ is log number of interstate highway lanes going through MSA c which proxies for its road infrastructure in year y. C_{jt} are city-specific time-varying controls from DT2011 that includes population, geography, census divisions, and socioeconomic characteristics. We include MSA (ζ_c) and year (ι_y) fixed effects.

We measure road traffic use with truck VKT from DT2011, since our focus is on commercial freight transport. To match truck traffic use (measured in vehicle-kilometers traveled), we measure rail traffic use in total railcar-kilometers-traveled for each city. Our detailed rail data also allow us to calculate rail traffic use by destination (incoming rail traffic use) and by origin (outgoing rail traffic use). We also observe rail car weight and construct an alternative measure of rail traffic use by weight in weight-

[22] Using the same three instruments, Duranton and Turner (2012) finds that a 10% increase in interstate highways increases city employment by about 1.5% between 1983 and 2003. With the same instruments, Duranton, Morrow and Turner (2014) finds that more highways lead to cities specializing in exporting goods with higher weight-to-value ratios.

[23] After the matching process, we have 221 MSAs. This is 7 fewer than DT2011 (see Section B.1.5 for more information). Despite the slight difference in observations, when we replicate the DT2011 results (Table A.2) our OLS and IV estimates have the same sign and are within one standard deviation of their results (Table 9, Columns (6) to (10) in DT2011).

kilometers-traveled. In order to measure relative changes between truck and rail use for U.S. cities in response to road infrastructure improvements, we take the ratio of rail traffic use (measured in railcar-kilometers or weight-kilometers) to road traffic (measured in truck vehicle-kilometers) for city c in year t $\left(\frac{\text{Rail Traffic Use}_{cy}}{\text{Road Traffic Use}_{cy}}\right)$.

OLS Results We begin by describing the road-specific and rail-specific responses, followed by the relative rail-to-road response. As mentioned earlier, a decrease in road cost will increase its traffic flows (Equation (3.8)). With our matched rail and truck data, we first show that we can replicate the significant positive road traffic use estimates from DT2011. Our OLS estimates (Columns (1) and (2), Table A.2) are within one standard error of their results (Columns (6) and (7), Table 9 of DT2011). Second, using our rail traffic use measure, we find a noisy, negative relationship with the decrease in road costs (Columns (1) and (2), Table A.4). These OLS coefficients are imprecisely estimated due to the two opposing substitution and complementarity forces shown in Equation (3.17) and discussed in Subsection 3.6: while a decrease in road costs raises rail costs relative to roads, thereby *reducing* rail traffic, it can also lower aggregate transport costs for the city and improve overall market access, thereby *increasing* demand for rail traffic.

Next, we study the relative modal split responses in rail and road traffic use. In Columns (1) and (2) of Table 1, we find that relative rail-to-truck traffic use has a negative and significant relationship with road infrastructure improvement in U.S. cities. This result is robust to including city-level and year-level fixed effects, as well as time-varying city-level controls such as socioeconomic characteristics and population. Note that this result, a coefficient of -1.4, is *not* mechanically driven by the inverse of the road traffic use estimate. In particular, this estimate is lower in absolute terms and differs by more than one standard error from the road traffic use OLS estimate of +1.6 from Table A.2 (Columns (1) and (2)). Further, this negative relationship is robust to measuring rail traffic use in terms of rail weight (Columns (1) and (2), Table A.5).

IV Strategy Since transport infrastructure and traffic use may be simultaneously determined, we require an instrumental variable (IV) approach to identify the causal effects of road infrastructure on rail traffic use and modal substitution. We employ the three instruments used in DT2011 to predict the stock of roads in U.S. cities and estimate the following two-stage least-squares IV regression:

$$\begin{aligned} \ln \text{Interstate Highway Lanes}_{cy} &= \eta_1 \ln \text{Instruments}_c + \kappa C_{cy} + \iota_y + \nu_{cy} \\ \ln Y_{cy} &= \eta_2 \ln \text{Interstate Highway Lanes}_{cy} + \phi C_{cy} + \iota_y + \mu_{cy} \end{aligned} \tag{4.2}$$

where $\ln Y_{cy}$ is log traffic use outcomes for MSA c in year y, $\ln \text{Instruments}_c$ is the set of three instruments discussed earlier, $\ln \text{Interstate Highway Lanes}_{cy}$ is log number of interstate highway lanes going through each city c, which proxies for the road infrastructure for that city in year y. C_{jt} are city-specific time-varying controls including population, physical geography, census divisions, and socioeconomic characteristics from DT2011, and ι_y is year fixed effects.

Table. 1. Elasticity of Modal Substitution

	(1)	(2)	(3)	(4)	(5)
Rail to Road Traffic Use	OLS	OLS	IV	IV	IV
Interstate Highway Lane KM	-1.432	-1.432	-0.867	-1.249	-1.099
	(0.195)	(0.196)	(0.376)	(0.388)	(0.364)
Population		-0.150	0.699	1.092	0.891
		(0.337)	(0.289)	(0.328)	(0.306)
Geography				✓	✓
Census Divisions				✓	✓
Socioeconomic Characteristics		✓			✓
MSA FE	✓	✓			
Year FE	✓	✓	✓	✓	✓
Observations	658	658	658	658	658
R-squared	0.88	0.88	-0.03	0.23	0.27
KP F-stat			14.48	10.76	10.04

Notes: Robust standard errors clustered by MSAs in parentheses. All variables are in logs. Rail traffic use, measured in railcar-kilometers, is constructed using confidential rail waybill data. Truck traffic use (in vehicle-kilometers) and other variables are from Duranton and Turner (2011). Instruments are 1835 exploration routes, 1898 railroad route kilometers, and 1947 planned interstate highways. Each regression includes 658 observations corresponding to 221 MSAs for each regression. See Table A.3 for first stage regressions.

The validity of the IV strategy requires that the instrument be uncorrelated with unobserved changes in both road and rail traffic use, including changes in the relative mode-specific market access (Appendix B.2), conditional on the control variables and fixed effects in Equation (4.2). As detailed in DT2011 and later papers, the first instrument, the 1947 highway plan, aimed to connect major 1940s population centers rather than anticipate future traffic or population growth. The second instrument, the 1898 railroad network, was developed by for-profit private companies operating during that time in a smaller agricultural economy, rather than to anticipate demand for transport 100 years later. The third instrument, the 1528-1850 exploration routes, were expeditions to find efficient paths between places, instead of anticipating transport demand in cities 150 years later. Additionally, controlling for historical population and physical geography helps address potential confounders in these instruments.

Results: Elasticity of Modal Substitution We first show that we can replicate the road traffic use IV results in DT2011 with our matched data set. Our IV estimates in Columns (3) and (4) (Table A.2) are within one standard error of their results (Columns (9) and (10), Table 9 of DT2011). Similar to them, we find that the IV estimates are slightly higher than the OLS estimates. Our results are also robust to including socioeconomic characteristics (Column (5), Table A.2). Next, we find that road infrastructure has a positive and noisy impact on rail traffic (Columns (3) to (5), Table A.4), suggesting that the complementarity effect dominates the substitution effect (Equation (3.17), Subsection 3.6).

Turning to the modal split responses, we find that road infrastructure improvements result in a larger increase in road traffic use relative to the rail increase—resulting in a decrease in the ratio of rail to road traffic use. A 1 percent increase in interstate highways causes a 0.9-1.2 percent decrease in rail to road

traffic use (Columns (3) to (5), Table 1). We use the estimate from the most conservative specification as our elasticity of modal substitution, $\eta = 1.099$ (Column (5), Table 1). Again, this result cannot be trivially attributed to the inverse of the road traffic use estimate: our estimates are lower in absolute terms and beyond one standard error from the road traffic use estimates, where a 1 percent increase in interstate highways causes a 1.7-2.1 percent increase in road traffic use (Columns (3) to (5), Table A.2).

To place our estimates in context, we next compare our modal substitution elasticity to the existing literature. Most of the economics literature focus on the substitution between ocean and air freight (Harrigan, 2010; Hummels and Schaur, 2013; Lugovskyy, Skiba and Terner, 2022; Tolva, 2025) or between bulk and container shipping within ocean freight (Coşar and Demir, 2018).[24] Compared to studies on ocean and air freight, our estimates are generally smaller in magnitude, likely due to the higher costs of air shipping leading to a higher elasticity of substitution.[25] In the transportation literature, our findings are broadly consistent with studies that focus on the substitution between rail and truck, with existing estimates falling within the confidence interval of our baseline estimate (Oum, 1979, 1989; Beuthe, Jourquin and Urbain, 2014).[26]

Robustness Checks We conduct several robustness checks of our elasticity of modal substitution. First, using an alternative measure of rail traffic by weight, we find estimates that retain the same sign and are within one standard error of our baseline (Columns (3) to (5), Table A.5). Second, since our rail traffic data is available at a more disaggregated level, we can examine incoming and outgoing rail traffic use separately. Using the ratio of directional rail traffic use to road traffic use, we find that these results again retain the same sign and are within one standard error of our baseline results, suggesting that both incoming and outgoing relative rail traffic use respond similarly to road infrastructure improvements (Columns (2) and (3) for incoming and outgoing traffic respectively, Table A.6). Third, we find similarly robust results combining both the alternative weight measure with directional rail traffic use (Columns (4) and (5) for incoming and outgoing traffic respectively, Table A.6).

Finally, while many of the 1898 railroads were abandoned and turned into roads (Duranton and Turner, 2011, 2012; Duranton, Morrow and Turner, 2014), there may still be concerns that the historical rail network remains correlated with the contemporary rail infrastructure. To address this concern, we further test the robustness of our results by excluding the 1898 railroads as an instrument. Under this specification, both elasticities using our carload and weight-based measures retain the same sign

[24] Instead of estimating an elasticity of modal substitution, Coşar and Demir (2018) estimates the cost structures of container and breakbulk shipping to quantify the impact of containerization in trade volumes.

[25] Hummels and Schaur (2013) estimates substitution elasticities of 2.7-6.5 for air versus ocean shipping, while Lugovskyy, Skiba and Terner (2022) reports an IV estimate of 10.3 for air versus containerized ocean shipping. Focusing on unit values, Harrigan (2010) finds that a 1 percent increase in value per weight raises the probability of air shipment by 0.2 percentage points.

[26] Differences in empirical strategies and context complicate direct comparisons. Beuthe, Jourquin and Urbain (2014) finds a rail ton-mile elasticity of 0.59 with respect to road cost reductions. Using 1970 Canadian freight flow data and a translog demand system, Oum (1979) estimates an rail-truck elasticity of 1.4-1.57 for eight selected commodities, and Oum (1989) estimates an elasticity of 1.19 for all goods.

and are within one confidence interval of our baseline results (Columns (6) and (7) for carload- and weight-based rail traffic use respectively, Table A.6).

4.2 Estimation of Intermodal Terminal Congestion

In this subsection, we first describe our approach to estimating the congestion elasticity λ_m at intermodal terminals by examining how ship loading and unloading times respond to port traffic. To account for the potential endogeneity of ship dwell times, we employ a shift-share instrument. We then describe our results and compare our estimates to the literature. Lastly, we conduct robustness checks to test the validity of our results.

Note that we apply λ_m to all intermodal terminals in our model, including inland rail terminals, even though it is based on ports. This is because the highly detailed minute-level AIS panel data, which enables us to precisely capture ship dwell times and include extensive fixed effects in our estimation, are not available at the same granularity for rail terminals. Despite this, by using a smaller railcar dwell time dataset, we find a significant and positive link between railcar dwell times and port traffic, highlighting the interconnectedness of transport modes and suggesting that port congestion can proxy for congestion at other terminals.

Empirical Strategy We measure port congestion by estimating the elasticity of ship dwell time with respect to port traffic. We estimate the following regression:

$$\ln \text{Ship Dwell Time}_{spdmy} = \beta_1 \ln \text{Port Traffic}_{pdmy} + \varphi_{sp} + \rho_{py} + \psi_{dmy} + \varepsilon_{spdmy} \tag{4.3}$$

where $\ln \text{Ship Dwell Time}_{spdmy}$ is the log number of hours ship s spent at port p on day of the week d in month m and year y, $\ln \text{Port Traffic}_{pdmy}$ is the log 28-day moving average amount of port traffic at port p ending on day d of month m and year y, φ_{spy} is ship-port fixed effects, ρ_{py} is port-year fixed effects and ψ_{dmy} is day-month-year fixed effects. The key parameter of interest, β_1, captures the elasticity of ship dwell times with respect to port traffic. Standard errors are two-way clustered at the ship and port levels.

The ship-port fixed effects control for fixed characteristics at the ship level, port level, and ship-port level. Fixed ship characteristics like ship sizes are important to control for since larger ships have larger capacities and naturally require more time to load and unload. Fixed port characteristics, such as the geographical proximity of ports to large trading partners or to multiple major cities, can also play a role in attracting more trade. This can, in turn, contribute to longer loading and unloading times (as seen in Figure 4b for LA and Long Beach). Fixed ship–port characteristics reflect time-invariant port-specific comparative advantages, including physical and geographic features such as deep natural harbors, that enable the accommodation of larger vessels and can ultimately increase dwell times. The port-year fixed effects control for time-varying port characteristics which includes potential technology or infrastructure

upgrades over time that may affect ship dwell times, such as larger cranes or expanded berths at ports over time to accommodate larger ships. Lastly, the day-month-year fixed effects control for aggregate events that impact all ships.

OLS Results We find that a 1 percent increase in port traffic is correlated with a statistically significant increase in ship dwell times by 0.1 percent (Column (2), Table 2). This elasticity is within one standard error of the baseline when estimated with ship fixed effects separately (Column (1), Table 2). West Coast ports, which are characterized by long histories of worker strikes, deeper harbors, and large volumes of U.S.-Asia trade (via LA and Long Beach), often experience longer ship dwell times. We show that the OLS coefficient for West Coast ports is indeed more than two times larger (Column (1), Table A.7). When including a pandemic period indicator (post-March 2020) to estimate separate coefficients during this time period, we find that the pre-March 2020 estimate is within one standard error of the baseline (Column (2), Table A.7), while the post-March 2020 estimates is slightly higher but still within one confidence interval of the baseline. Using shorter periods for the moving average calculation, we find that the ship dwell times are still statistically significantly correlated with port traffic, but the magnitudes of the coefficients are smaller (Table A.8).

IV Strategy To establish the causal impact of port traffic on ship dwell times, we require a demand shifter for port traffic that is uncorrelated with unobserved ship dwell time determinants ε_{spdmy}. Our IV strategy leverages aggregate origin country- and product-level compositional changes for imports into top U.S. ports to predict the transport services demand at each port:

$$\text{Port Trade Exposure}_{pmy} \equiv \sum_{O}\sum_{N} X_{on\backslash p,my} \times \omega_{onp,2003} \tag{4.4}$$

where the shift-share IV is the weighted sum of top 30 U.S. ports' imports of product n from origin o excluding port p at month m and year y. The first term in Equation (4.4) is the contemporaneous shift in overall imports into top U.S. ports excluding its own port p, capturing month-to-month fluctuations in demand for port services across origin-product pairs. The second term is a port-specific lagged weight, calculated from import shares at least 13 years prior at the start of the available sample period (year 2003). The weights sum to one across products and origins for each port-year observation. The Census Bureau monthly trade data allow us to match this IV with our 28-day moving average of port traffic.

For this IV strategy to be valid, the port trade exposure measure has to be generally uncorrelated with unobserved determinants of dwell times for ship servicing these ports after controlling for fixed ship-port characteristics, time-varying port characteristics, and aggregate time-varying events that impact all ships. Following recent best practices in shift-share IV designs (Borusyak, Hull and Jaravel, 2025), we report both the effective sample size and the distribution of the instrument. The effective sample size of the instrument's exposure share is relatively high at 204.8, suggesting that our instrument is not overly influenced by a small number of dominant shocks. Furthermore, the largest exposure share accounts for

Table. 2. Congestion Elasticity of Port Traffic with respect to Ship Dwell Times

	(1) OLS	(2) OLS	(3) First-Stage	(4) IV	(5) First-Stage	(6) IV
Port Traffic	0.093	0.100		0.264		0.236
	(0.011)	(0.010)		(0.095)		(0.091)
Port Trade Exposure			0.229		0.228	
			(0.013)		(0.014)	
Day-Month-Year FE	✓	✓	✓	✓	✓	✓
Port-Year FE	✓	✓	✓	✓	✓	✓
Ship-Port FE		✓			✓	✓
Ship FE	✓		✓	✓		
Observations	90516	90516	90516	90516	90516	90516
First Stage KP-F				299.03		274.05

Notes: Robust standard errors in parentheses are two-way clustered at the ship and port levels. All variables are in logs. Port traffic is the 28-day moving average of total daily net tonnage at the port. Weighted by ship net tonnage.

only 2.9% of the total, suggesting a reasonable degree of dispersion.[27] Additionally, we establish that our IV strategy has a strong first stage: the port trade exposure measure has a statistically significant and positive relationship with port traffic (Column (5) and (6), Table 2), with Kleibergen–Paap F-statistics exceeding the Staiger and Stock (1997) threshold.

To further support the validity of our identification strategy, we conduct a balance test to show the absence of correlation between the instrument and an approximation that is correlated with of unobserved ship dwell times determinants. One potential concern is that that time-varying port-ship interactions, which is outside the scope of our fixed effects, might affect our results. Specifically, busier ports may invest in infrastructure, such as increasing the number of container cranes, to accommodate growing containership traffic and reduce dwell times of these ships. To proxy for this type of time-varying port-specific investments targeted at containerships, we use data from the Bureau of Transportation Statistics (BTS) on the annual number of container cranes which is available for the top 25 U.S. container ports for a subset of years. We residualize both the number of container cranes and the port trade exposure instrument with respect to port and time fixed effects and find a noisy correlation between the two (Figure A.5). This lack of correlation suggests that our instrument is unlikely to be systematically related to endogenous, time-varying port investments aimed at particular ships.

Results: Elasticity of Intermodal Terminal Congestion We find that a one percent increase in port traffic results in a statistically significant 0.24 percent increase in ship dwell time (Column (6), Table 2). This result is within one standard error of the specification that separately includes ship fixed effects (Column (4), Table 2). This congestion elasticity forms the basis for congestion elasticity at terminals: parameter λ_m (Equation (3.13)). Since this elasticity is measured in time while Equation (3.13) is

[27] Across product and origin countries, the effective sample size is 87.8, with the largest weight at 5.6%.

expressed in cost terms, we convert this time elasticity into a cost measure using the ad valorem tariff estimate for an additional day in transit from Hummels and Schaur (2013), resulting in λ_m=0.096.

Without directly comparable elasticities on intermodal terminal congestion, we compare our estimate with two strands of prior literature. First is the literature on processing times at ports. Using detailed Peruvian data, Carballo et al. (2021) estimates an import processing cost elasticity of 0.06, which is within the confidence interval of our baseline estimate.[28] Second, we compare our estimates to congestion elasticities from the urban literature which focuses on road congestion in cities. Existing U.S. estimates range from 0.03-0.11, also falling within one confidence interval of our baseline estimate (Couture, Duranton and Turner, 2018; Allen and Arkolakis, 2022; Akbar et al., 2023; Duranton and Puga, 2023).[29]

Robustness Checks As noted earlier, West Coast ports typically have longer dwell times due to their location and history of labor strikes. Indeed, we estaimte a higher congestion elasticity for these ports (Column (3), Table A.7) that is still within one confidence interval of our baseline result. As additional checks, we examine the periods before and after the pandemic separately. While the post-pandemic period has a higher congestion elasticity (as we would expect), both elasticities are within one standard error of the baseline (Columns (4) and (5) respectively, Table A.7).

Although our congestion elasticity λ_m is applied to all intermodal terminals within the multimodal network, it is estimated solely from AIS vessel traffic data at ports. This is due to the exceptional level of detail provided by these panel data, which allows us to capture ship dwell times with precision and incorporate an extensive range of fixed effects. To the best of our knowledge, comparable rail station dwell times data with the same level of granularity are not available. However, by using a smaller sample of rail dwell times matched to port traffic data for nearby rail stations, we find a statistically significant and positive relationship between port traffic and rail dwell times at these rail stations (Table A.9, Appendix Section B.5).[30] This finding highlights how congestion at ports likely leads to delays at other key intermodal terminals due to the interconnectedness of transport modes within the multimodal network. Consequently, port congestion can serve as a proxy for understanding congestion impacts at intermodal terminals across the broader multimodal network.

[28] Carballo et al. (2021) estimates that a 1 percent increase in median import processing time reduces import values by 0.24%, giving a processing cost elasticity of 0.06 with a demand elasticity of 4. With export processing data on 98 countries, Djankov, Freund and Pham (2010) finds that a 1 percent increase in delivery time from factory to port reduces trade by 0.4%.

[29] We focus on U.S. estimates for comparison purposes. Couture, Duranton and Turner (2018) estimate a congestion elasticity of 0.11 associated with U.S. metro area population. Duranton and Puga (2023) estimates an elasticity of 0.04 using residualized city speed for 180 U.S. metropolitan areas. Using a global dataset on vehicle travel speeds, Akbar et al. (2023) estimate a congestion elasticity with respective to income of 0.032 and provides an in-depth discussion of on existing congestion elasticity estimates. AA2022 estimate a road congestion elasticity of 0.092.

[30] See Appendix Section B.1.6 for further details on the rail station dwell times data.

5 Calibration and Validation

To conduct counterfactuals using our framework, we begin by presenting two propositions that in tandem define a system of equations for changes within the multimodal transport network. Next, we discuss how we use detailed traffic and geography data to construct the US multimodal transport network and the calibrate key parameters in the model. We then conduct three validity checks of our framework by (1) comparing observed and model-predicted mode-specific trade flows, (2) examining how both observed and predicted trade flows vary with distance to assess gravity model implications, and (3) comparing the distribution of observed and predicted trade across transport modes over distance.

5.1 Counterfactual Equilibrium

Our goal is to evaluate the welfare impact of infrastructure investments or disruptions in a setting where agents make sophisticated routing and mode choices within a multimodal transport network. To accomplish this, we derive a system of equations that determines the counterfactual equilibrium for arbitrary mode-specific cost changes within the multimodal network while also allowing for a rich characterization of congestion forces. We first extend the equilibrium system in Equation (3.4) to allow for mode choice (as in Section 3.4) and for congestion (as in Section 3.5). Following Dekle, Eaton and Kortum (2008), we then employ 'Hat Algebra' to express the equilibrium in terms of changes in the endogenous variables. We denote the changes in variables with hats, $\hat{\gamma}_i \equiv \frac{\gamma_i'}{\gamma_i}$. The structure of this model allows for a convenient separation of the link- or edge-specific transport equilibrium *given* market access terms and the determination of the aggregate equilibrium *given* (aggregate) transport costs.

We present the counterfactual equilibrium in two parts: one characterizing the aggregate equilibrium and the other describing the edge-level transport equilibrium subroutine. Detailed derivations for both are provided in Appendix A.10.

Proposition 1. *Consider an economy in equilibrium with a primary transport network,* $\overline{\mathbf{T}} \equiv \left[\bar{t}_{kl,1}\right]$*, a secondary transport network (m),* $\overline{\mathbf{T}}_m \equiv \left[\tau_{kl,m}\right]$*, and an intermodal terminal network connecting both primary and secondary networks (m),* $\overline{\mathbf{S}}_m \equiv \left[\bar{s}_{kk,m}\right]$*. Consider any changes either in the primary infrastructure network,* $\hat{\bar{t}}_{kl,1}$*, in the intermodal terminal network,* $\hat{\bar{s}}_{kk,m}$*, or in the secondary transport network,* $\hat{\tau}_{kk,m}$*. Given observed aggregate and mode-specific traffic flows* $\left(\Xi_{ik}, \Xi_{ik,m}\right)$*, economic activity in the geography* (Y_i, E_j)*, and parameters* $\{\alpha, \beta, \theta, \lambda_1, \lambda_m, \eta\}$*, the equilibrium change in economic outcomes* $\left(\hat{y}_i, \hat{l}_i, \hat{\chi}\right)$ *is the solution to the following :*

1. ***Aggregate Equilibrium.*** *Changes in incoming and outgoing market access terms* $\left\{\hat{\Pi}_i, \hat{P}_j\right\}$ *given by:*

$$\begin{aligned}\hat{P}_j^{-\theta} &= \left(\frac{\gamma_j}{\gamma_j + \sum_{k\in\mathscr{B}(j)}\Xi_{kj}}\right)\frac{\hat{\gamma}_j}{\hat{\Pi}_i^{-\theta}} + \sum_{k\in\mathscr{B}(j)}\left(\frac{\Xi_{kj}}{\gamma_i + \sum_{k\in\mathscr{B}(j)}\Xi_{kj}}\right)\hat{t}_{kj}^{-\theta}\hat{P}_k^{-\theta} \\ \hat{\Pi}_i^{-\theta} &= \left(\frac{\delta_i}{\delta_i + \sum_{k\in\mathscr{F}(i)}\Xi_{ik}}\right)\frac{\hat{\delta}_i}{\hat{P}_i^{-\theta}} + \sum_{k\in\mathscr{F}(i)}\left(\frac{\Xi_{ik}}{\delta_i + \sum_{k\in\mathscr{F}(i)}\Xi_{ik}}\right)\hat{t}_{ik}^{-\theta}\hat{\Pi}_k^{-\theta}\end{aligned} \tag{5.1}$$

where $\hat{t}_{ki}^{-\theta} = F\left(\left\{P_j, \Pi_i\right\}\right)$ *is the change in endogenous transport costs derived from the transport equilibrium in Proposition (1.2),* $\delta_i = \gamma_i = y_i$ *represents the share of income in location i, and changes in the market access terms* $\hat{P}_j$ *and* $\hat{\Pi}_i$ *are equal to changes in income and labor densities,* $\hat{P}_j = \hat{y}_j\hat{l}_j^{\beta-1}\hat{W}^{-1}$ *and* $\hat{\Pi}_i = \hat{l}_i^{\alpha+1}\hat{y}_i^{-\frac{\theta+1}{\theta}}$ *where* y_i *is the share of world income in i,* $y_i \equiv \frac{Y_i}{Y^W}$*, and* l_i *is the share of total labor in i,* $l_i \equiv \frac{L_i}{L^W}$*.*

2. ***Transport Equilibrium.*** *Given* $\left\{\hat{P}_i, \hat{\Pi}_i, \hat{\gamma}, \hat{\delta}\right\}$*, the equilibrium change in endogenous aggregate transport costs* $\left(\hat{t}_{ik}^{-\theta}\right)$ *is the solution to the edge-level transport equilibrium and given by:*

$$\hat{t}_{ik}^{-\theta} = \left(\sum_{m\in\mathscr{M}(i,k)}\frac{\Xi_{ik,m}}{\Xi_{ik}}\hat{t}_{ik,m}^{-\eta}\right)^{\frac{\theta}{\eta}} \tag{5.2}$$

and changes in the endogenous transport cost on the primary mode and any alternative mode are given by, $\hat{t}_{kl,1} = \hat{\tilde{t}}_{kl,0}^{\frac{1}{1\cdot\lambda_1}}\hat{t}_{kl}^{\frac{\lambda_1(\eta-\theta)}{1+\eta\lambda_1}}\hat{P}_k^{\frac{-\theta\lambda_1}{1+\gamma\lambda_1}}\hat{\Pi}_l^{\frac{-\theta\lambda_1}{1\cdot r\lambda_1}}$ *and* $\hat{t}_{kl,m} = \hat{\tilde{s}}_{kk,m}\hat{\tau}_{kl,m}\hat{\tilde{s}}_{ll,m}\hat{P}_k^{-\theta\lambda_m}\hat{\Pi}_{k,m}^{-\theta\lambda_m}\hat{\Pi}_l^{-\theta\lambda_m}\hat{P}_{l,m}^{-\theta\lambda_m}$ *where the mode specific market access terms are defined as,* $\hat{P}_{l,m}^{-\theta} \equiv \sum_{k\in B(l)}\frac{\Xi_{kl,m}}{\sum_{k\in\mathscr{B}(l)}\Xi_{kl,m}}\hat{t}_{kl,m}^{-\eta}\hat{t}_{kl}^{\eta-\theta}\hat{P}_k^{-\theta}$ *and* $\hat{\Pi}_{k,m}^{-\theta} \equiv \sum_{l\in\mathscr{F}(k)}\frac{\Xi_{kl,m}}{\sum_{l\in\mathscr{F}(k)}\Xi_{kl,m}}\hat{t}_{kl,m}^{-\eta}\hat{t}_{kl}^{\eta-\theta}\hat{\Pi}_l^{-\theta}$*.*

Proposition 1 constructs the aggregate counterfactual spatial equilibrium solving for market access terms while simultaneously solving for the modal traffic assignment. The algorithm works like a nested fixed point problem. The first part of Proposition (1.1) solves for the aggregate equilibrium, while the second part represents a nested subroutine which solves for the modal transport flows conditional on a guess for the market access terms (Proposition (1.2)). A computational benefit arises because—given a candidate solution for the aggregate equilibrium—the transport equilibrium is entirely disentangled from the determination of the aggregate spatial equilibrium, while the aggregate equilibrium inherits the computational properties in terms of speed and scalability of a (gravity) spatial equilibrium system.

From a practitioner's point of view, the proposition implies that given observed traffic flows on the primary network, bilateral flows on the secondary network[31], $\left(\Xi_{ik}, \Xi_{ik,m}\right)$ as well as knowledge of the model parameters, $\{\alpha, \beta, \theta, \lambda_1, \lambda_m, \eta\}$, we can employ the model to evaluate infrastructure improvements along the primary network or at terminal stations, thus improving the connectedness of the primary and secondary transport network. We further extend the model to allow for a partitioned

[31]Notice the slight abuse of notation here. While $\Xi_{ij,1}$ refers to the link- or edge-specific traffic along the primary network, $\Xi_{ij,m}$ instead refers to modal flows between i and j along the secondary networks and is therefore not necessarily edge-specific. However, $\Xi_{ij,m}$ summarizes modal traffic in the sense that it refers to any flows between i and j no matter their origin or destination on the primary network. This is convenient—as we will argue below—since this is the data moment that is directly observed in the rail traffic data.

set of domestic and foreign locations with international trade between a subset of domestic nodes and foreign locations, as presented in Appendix A.12.

Relative to the existing literature, the proposition provides a straightforward extension of Equations (28) and (29) in AA2022. The only difference is the presence of a nested subroutine, which, given a candidate solution for spatial equilibrium $\{P_i, \Pi_i\}$, solves the transport equilibrium at the edge-level (i.e. the set of equations defined by Equation (5.2)). This implicitly determines the modal choice, taking into account endogenous congestion, and generates an endogenous aggregate transport cost.[32] This adjustment adds a novel channel by which to evaluate infrastructure improvements. In this setting, the impact of improving transport infrastructure has the same direct and general equilibrium effect as in AA2022 where route choice is impacted, congestion can be alleviated, and input and output markets can adjust, all of which collectively add up to welfare gains. In our setting, we also feature a direct interplay between the primary and secondary networks. Mode-specific infrastructure investments can lead to modal diversion and thus alleviate congestion on the alternative transport network. The extent to which this might occur depends on the cross-sectional variation in the access to the secondary transport system, which is reflected by variations in the weights on the final term across space.

5.2 Model Calibration to the US Multimodal Network

In order to calibrate our framework to the US multimodal network, we require data on income and population as well as geo-spatial and traffic data on roads, rail, waterways, ports, and inland intermodal terminals. We utilize the existing income, population, as well as road geo-spatial and traffic data from AA2022. We then introduce additional geo-spatial and traffic data for rail, waterways, ports, and intermodal terminals, as discussed in Section 2.1. To allow for international trade, we obtain Census Bureau data on flows from international locations to domestic ports.[33] Our resultant sparse graph representation of the multimodal transport network collapses the high-dimensional geo-spatial information from each mode-specific shapefile and retains nodes at endpoints or intersections. Cities containing at least 10,000 people (CBSAs) are also represented by a single node within the network. The resulting graph consists of 228 nodes and 704 edges.

We now discuss our choice of model parameters $\{\alpha, \beta, \theta, \lambda_1, \lambda_m, \eta\}$. For the local productivity

[32] We can *explicitly* characterize the change in the equilibrium traffic flows along the primary and secondary transport system. Given the equilibrium changes in market access terms $\left(\hat{P}_i, \hat{\Pi}_i, \right)$, as well as the equilibrium changes of mode and aggregate transport costs, $\left(\hat{t}_{ik}, \hat{t}_{ik,m}, \right)$, the resulting change in the traffic flows can be computed using the following formulae:

$$\begin{aligned} \hat{\Xi}_{ik} &= \hat{t}_{ik}^{-\theta} \times \hat{P}_i^{-\theta} \times \hat{\Pi}_k^{-\theta} \\ \hat{\Xi}_{ik,m} &= t_{ik,m}^{-\eta} \times \hat{t}_{ik}^{\theta-\eta} \times \hat{P}_i^{-\theta} \times \hat{\Pi}_k^{-\theta} \end{aligned} \tag{5.3}$$

This allows us to account for the changes in observed traffic flows. This can be a convenient tool to analyze the implied traffic and environmental impacts of infrastructure investments.

[33] To limit the dimensionality of the simulation, we group the flows into five geographic regions: Africa, Asia, Australia and Oceania, Europe, North America and South/Central America.

spillover α, local amenity spillover β, and shape parameter θ, we follow the literature and set $\theta = 8$, $\alpha = .1$ and $\beta = -.3$ AA2022. For the road congestion elasticity, we use $\lambda_1 = 0.092$ following AA2022 who estimates this parameter by regressing the observed speed on individual highway segments against a measure of instrumented traffic.

The last two parameters come from our own estimates. For the intermodal terminal congestion parameter λ_m, we first determine the impact of port traffic on the observed duration a ship spends at port using an instrument which measures the trade exposure at these ports (Table 2, Section 4.2). We use $\lambda_m = 0.096$ after converting this time elasticity into a cost measure, relying on the ad valorem tariff estimate on an additional day in transit from Hummels and Schaur (2013). For the modal elasticity of substitution η, we estimate the impact of road infrastructure improvements on rail relative to road traffic flows using the instruments from Duranton and Turner (2011). We use $\eta = 1.099$ (Table 1, Section 4.1). The non-overlapping confidence intervals for η and θ highlight a significant difference between them, supporting the validity of our modeling approach which allows for two distinct elasticities to govern transport mode and route-sourcing decisions separately. As discussed earlier in Section 3, this also suggests that a simple extension of the AA2022 to allow for mode-specific subnetworks is insufficient to accurately capture both decisions (footnote 8).

5.3 Predicted versus Observed Trade Flows by Transport Mode and Distance

We conduct three comparisons to assess the ability of our model to capture actual observed trade flows for different transport modes. First, we compare our model predictions for mode-specific trade flows between CBSAs against commensurate data from the Commodity Flow Survey (CFS). Second, we examining its gravity model implications by comparing the relationship between observed trade flows and distance to the relationship between the predicted trade flows and distance. Third, we compare our model predictions for the distribution of freight transported over multiple modes over distance against observed data from the Department of Transportation.

Mode-specific trade flows In order to assess the ability of our model to capture actual observed trade flows for different transport modes, we compare our model predictions for mode-specific trade flows between CBSAs against commensurate data from the Commodity Flow Survey (CFS). The model-predicted mode-specific trade flows are calculated using Equation (A.87) and observed mode-specific traffic flow data from the CFS.[34] We include all three transport modes—truck, rail, and barges. Figure 6a shows the scatter plot between the observed and predicted mode-specific trade flows, conditional on origin and destination fixed effects. This allows for a pair-wise comparison between the observed and predicted mode-specific flows at the origin-destination-level, independent of origin or destination characteristics. We find a strong positive correlation of 0.75 with all three modes, indicating that the

[34] See Online Appendix A.11 for detailed derivations of the model predicted mode-specific trade flows.

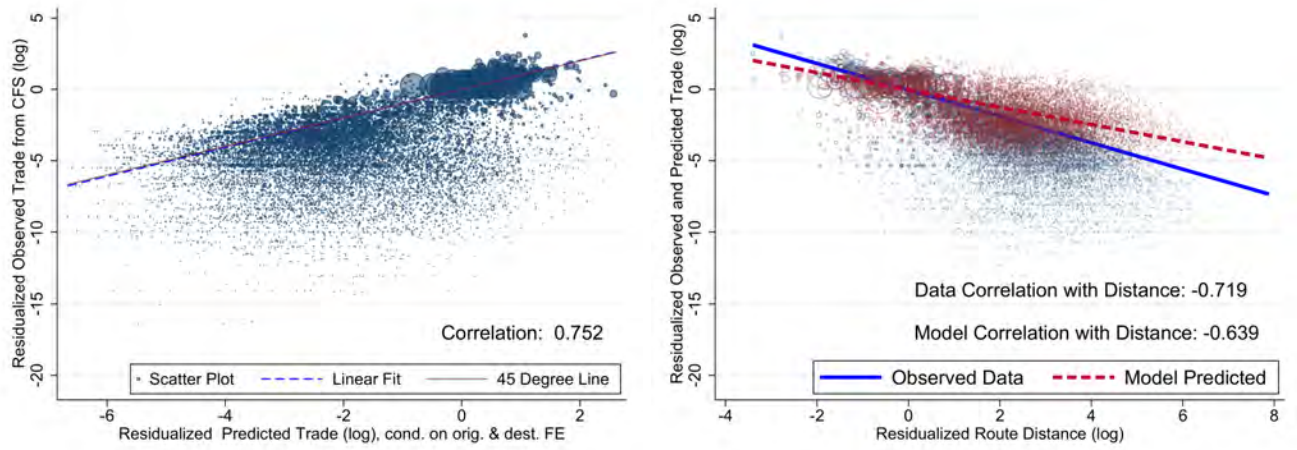
model predictions reflect observed mode-specific trade flows. These results are also an improvement in terms of model fit relative to previous literature focusing on one mode. As a comparison point, AA2022 finds a strong positive correlation of 0.60 using truck flows. Column (1) of Table A.10 reports the univariate regression outcome between the residualized observed and predicted trade flows, weighted by trade weight in tons. We find a significantly positive relationship, with a coefficient of 1 within one standard error.

Gravity model implications Additionally, we can evaluate the validity of our model by examining its gravity model implications. The standard gravity model implies a negative log-linear relationship between trade and trade costs at the bilateral level, both internationally between countries (Head and Mayer, 2014) and domestically between cities within a country (Allen and Arkolakis, 2014; Fajgelbaum and Schaal, 2020). Following previous literature, we compare the relationship between observed trade flows and distance to the relationship between the predicted trade flows and distance, similarly conditioned on origin and destination fixed effects as above. Figure 6b find similar strong negative correlations between the two, strong negative correlations of 0.64 in the model and 0.72 in the data. We find significantly negative relationships in linear regressions for both, with a negative coefficient of 0.61 in the model and 0.93 in the data. The similarity between the predicted and observed relationships with distance strengthens the validity of our model, suggesting that it provides reasonable approximations for the distribution of trade flows.

Looking into specific transport modes, we find strong positive correlations of 0.81 and 0.56 between the residualized observed and predicted trade flows for trucks and multiple modes respectively. Column (2) of Table A.10 reports the regression outcome between the residualized observed and predicted trade flows for the three transport modes, weighted by trade weight in tons. We find a significantly positive relationship that is close to 1 for all three modes, with the rail coefficient slightly higher, suggesting that observed rail flows exceed model predictions. This may be due to the CFS data capturing flows for all commodities, while our model only predicts containerized flows. From our confidential waybill data, containerized trade accounts for 43 percent of all rail car loads in 2012.

Trade distribution over distance Third, we assess our model predictions for the distribution of freight across transport modes over distance by comparing them with observed data from the Department of Transportation (Figure 1). Our model closely matches the initially low multimodal share at short distances (below 100 miles), and captures its steady increase as distance grows, including its consistently high shares above 70% for long distances above 1,000 miles. The model does predict a slightly earlier shift towards multimodal use relative to the data in the intermediate distance range (100-999 miles). Overall, our model exhibits a strong correlation of 0.97 between predicted and observed multimodal freight shares over distance (Figure 7). Notably, our model achieves this fit using a single elasticity of modal substitution parameter, without allowing substitution patterns to vary with distance.

Figure 6. Model Fit: Predicted and Observed Trade Flows



(a) Predicted and Observed Bilateral Trade Flows **(b)** Trade Flows versus Distance

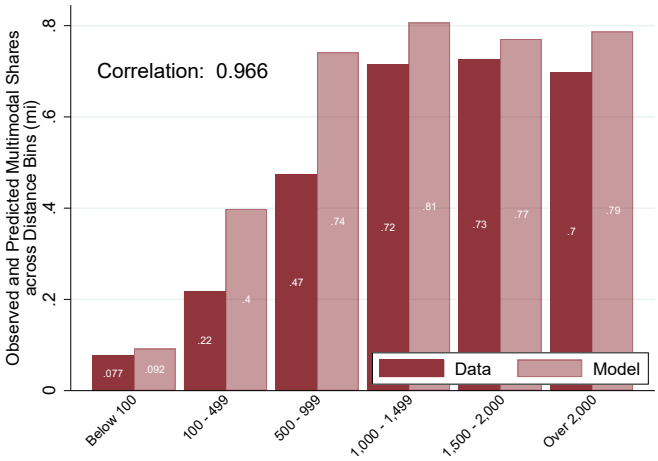
Notes: Figure 6a visualizes the comparison between observed bilateral origin to destination mode-specific trade flows with the mode-specific trade flows predicted by the multimodal economic geography model based on observed traffic data along the transport network. The observed mode-specific trade flows between metropolitan areas on the y-axis are obtained from the 2012 Commodity Flow Survey (CFS) data while the predicted mode-specific trade flows are on the x-axis. Both the observed and predicted trade flow measures are in logs and residualized using origin and destination fixed effects. This allows for the comparison to come from similarities at the origin-destination pair-level. Figure 6b visualizes the observed and predicted bilateral trade flows against distance, all residualized using origin and destination fixed effects. In both panels, circle sizes indicate trade weight in tons.

Implied Modal Wedges. To summarize the gap between observed modal cost advantages and observed modal adoption, we construct an implied modal adoption wedge from the estimated elasticity $\hat{\eta}$. For each non-road mode m, the observed road-relative shipment ratio pins down the generalized-cost ratio required by the model; comparing that revealed ratio with the calibrated per-TEU-mile cost ratio yields a multiplicative wedge Λ_m, interpreted as the extra generalized friction that must be loaded onto mode m for the model to match observed use. This object is informative because it is imputed from behavior rather than observed directly: it summarizes the combined effect of transfer costs, schedule inflexibility, reliability concerns, coordination frictions, access limitations, and other non-line-haul barriers that can prevent modal adoption even when engineering transport costs favor rail or water. At our baseline $\hat{\eta} = 1.099$, the average local common-support wedge is 14.17 for rail, 30.98 for domestic water, and 28.30 for foreign water. Translating these into dollar-equivalent values at the benchmark haul distance $D = 537.3$ miles yields an average-local rail equivalent of $\$5,585$ per TEU, with an exact national counterpart of $\$2,013$. We therefore interpret Λ as a reduced-form revealed measure of modal adoption frictions, rather than as a literal terminal fee or switching charge. Detailed derivations in A.9.

6 Network Improvements, Decompositions, and Policy Scenarios

In this section, we apply our framework to evaluate the economic and environmental impacts of targeted infrastructure investments and policy-relevant scenarios, taking the multimodal nature of the U.S.

Figure 7. Model Fit: Predicted and Observed Transport Mode Distribution over Distance



Notes: This figure presents a comparison of observed shares of freight transported multimodally across various distances (in meters), from the Department of Transportation (DoT), with model predicted multimodal shares.

domestic transport system into account.[35] Our main application compares the welfare impact of investing in different terminals across the U.S., improving the intermodal integration of the multimodal transport network. Next, to illustrate the relative importance and contribution of the different channels in our model, we conduct four decompositions using counterfactual improvements to the U.S. highway network as a benchmark. Additionally, we quantify the welfare changes and environmental impacts of three policy-relevant scenarios: (1) the impact of losing rail network access, motivated by the averted rail strike in the fall of 2022, (2) the impact of repealing the Merchant Marine Act of 1920 (also known as the Jones Act), which restricts ships transporting goods between U.S. ports to those owned, crewed, and built by U.S. citizens, and (3) the impact of disruptions to the Panama Canal due to drought conditions.

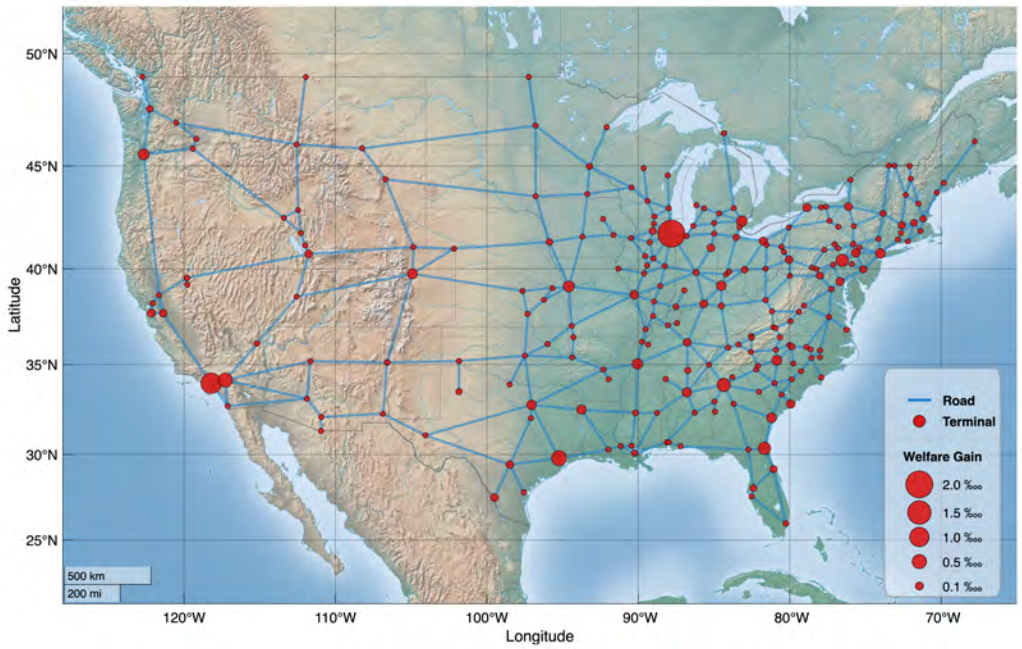
6.1 Gains from Intermodal Terminal Improvements

While previous papers have focused on estimating the welfare effects of improving individual segments of the U.S. road network or individual ports, less is known about the welfare impact of improving the degree to which the U.S. multimodal transport network is interconnected. In order to evaluate this, we will use the counterfactual equations of Proposition (1) to estimate the aggregate welfare impact ($\hat{W} = \hat{\chi}^{-\frac{1}{\theta}}$) of a small (1%) improvement to the switching cost, $s_{ii,m}$, at each of the 228 intermodal nodes across the network, given the calibrated parameters and observed traffic on road, rail, waterways, rail terminals, and ports.

Our results from these 228 counterfactual simulations are visualized in Figure 8 and the top 10

[35] We explicitly embrace a computational approach in this section, employing the 'hat algebra' developed in Section 5.1. Alternatively, one could potentially derive sufficient statistics that characterize welfare elasticities in terms of traffic flows as in Office of the State Auditor (2019) and in principle the recursive formulation of the equilibrium condition (Equation (3.4)) might be a convenient aid in doing so as highlighted in our discussion in Subsection 3.17. However, the presence of congestion elasticities on the primary mode and intermodal terminals makes this approach less desirable.

Figure 8. Welfare Benefits of Improving Intermodal Terminals



Notes: This figure visualizes the welfare impact of lowering the transshipment cost in each intermodal terminal node by 1 percent. The network includes 228 nodes in total, each corresponding to a separate counterfactual. The red circle size is proportional to the magnitude of welfare gains, with larger circles indicating larger welfare effects. The blue lines indicate the graph representation of the primary road network. State boundaries are included. See Table 3 for details on the top 10 terminals by welfare impact and Table A.11 for the top 30.

terminals with the highest impact are listed in Table 3. For the welfare benefits, we convert the welfare elasticity into a dollar amount using a compensating variation approach by calculating the necessary increase, in millions of chained 2012 U.S. dollars, in annual U.S. real GDP (valued at $19 trillion) to achieve the same welfare improvement that we estimate. Many of the most impactful terminals are located in the center of the United States, including Chicago, Atlanta, and Kansas City, highlighting the role of multimodal network transporting goods from coastal regions to the interior. Chicago, the top-ranked terminal, plays a central role as an intermodal hub for the Midwest, as reflected in its highest rail throughput (Table 3, Column (3)). Unsurprisingly, several of the high-ranking intermodal terminals are also located in populous coastal cities like Los Angeles (which also includes Long Beach) and Houston, which rank second and third respectively. Several locations, like Riverside CA and Harrisburg PA, do not particularly large throughput flows, but their importance comes from their strategic positions within the multimodal system, linking nearby major cities like Los Angeles and Philadelphia to the broader network (Figure 8). The implied welfare benefit of lowering the switching costs in the most central nodes could represent a welfare gain equivalent to increasing U.S. GDP between $463-3,851 million (in 2012 USD, Table 3, Column (6)).

Our results are complementary to Brancaccio, Kalouptsidi and Papageorgiou (2024), which analyzes U.S. port infrastructure in the context of dry bulk commodities like grain and coal and identified that the East Coast and Gulf ports generate the highest welfare returns from increases in effective port

Table. 3. Ranking: Welfare Benefits of Improving Intermodal Terminals

	(1) CBSA Name	(2) Population	(3) Throughput	(4) ROI	(5) Benefit ($m)	(6) Cost ($m)
1	Chicago	9368268	10368684	3.730	3851	814
2	Los Angeles	9639715	6836640	3.039	2168	537
3	Houston	3133212	630300	24.782	1086	42
4	Riverside	2173638	761760	10.999	942	79
5	Atlanta	1627623	1830840	4.643	811	144
6	Harrisburg	655561	793920	10.194	698	62
7	Jacksonville	936317	797880	17.252	689	38
8	Kansas City	1767872	1088760	5.335	510	81
9	Portland	1641801	424296	16.349	492	28
10	Detroit	2732964	557760	11.799	463	36

Notes: The table shows the ten terminals where a one percent reduction of the transshipment cost generates the highest benefit. Column (1) indicates the core based statistical areas (CBSA) name of the node, which includes both metropolitan and micropolitan areas. The terminal's population and rail throughput in TEUs are reported in Columns (2) and (3) respectively. Column (4) shows the imputed return on investment (ROI), Column (5) calculates how much 2012 U.S. GDP would need to increase in order to match the overall welfare gain, while Column (6) presents the required cost of making this one percent cost decrease. For an extended version including the top 30 terminals see Table A.11.

capacity. Even though our focus is on containerized trade at intermodal terminals which includes both ports and inland terminals, we similarly find large welfare gains from improvements at major ports like Houston and Portland which handles large volumes of both container and dry bulk. Additionally, we find significant welfare returns from investments in interior intermodal terminals, such as Chicago and Atlanta. Our results are also complementary to AA2022 which identifies gains from highway link improvements in short coastal segments linking densely populated areas (like Boston-Philadelphia and Los Angeles-San Diego) and trade thoroughfares via Indiana. While our results also identify gains for large cities like Los Angeles, our improvement gains are mostly in the central United States. This distinction highlights the importance of multimodal transport network in facilitating long-distance trade and linking coastal to interior regions (as indicated in Figure 1). These interior hub terminals are highly central to the transportation system and represent potential bottlenecks in the multimodal network.

Additionally, we combine our welfare benefits with cost assessments in order to compute the return on investment (ROI) for each of these terminal investments. To operationalize this, we utilize equation (42) in AA2022 adapted for terminal operations to determine the amount of container volume required to achieve a one percent reduction in switching cost at these terminals.[36] Next, we calculate the number of gantry cranes in container terminals required to achieve this higher container volume using industry estimates. Since these terminals require land, each location will have a different construction cost based on local property prices which we obtain from Albouy, Ehrlich and Shin (2018). Assuming a 20-year depreciation schedule, a 5% annual maintenance cost, and a 3% borrowing cost per existing literature, we estimate that 10% of the construction cost is accrued on an annual basis. We estimate high terminal

[36]This implies a (log) proportional relationship between number of terminals and switching costs, i.e. $\hat{s}_{kk,m} = \hat{terminals}_{kk,m}^{-\lambda_m}$. where $\hat{s}_{kk,m}$ denotes the mode specific (log) change in the switching cost and $\hat{terminals}_{kk,m}$ denotes the necessary (log) change in terminal capacity.

improvement costs for populous cities with higher land prices like Chicago and Los Angeles (Table 3, Column (7)), translating into low ROIs for these locations (Table 3, Column (5)). On the other hand, we estimate lower construction costs for locations with relatively cheaper land prices like Houston, Jacksonville FL, and Portland OR which translates into high ROIs for these locations.

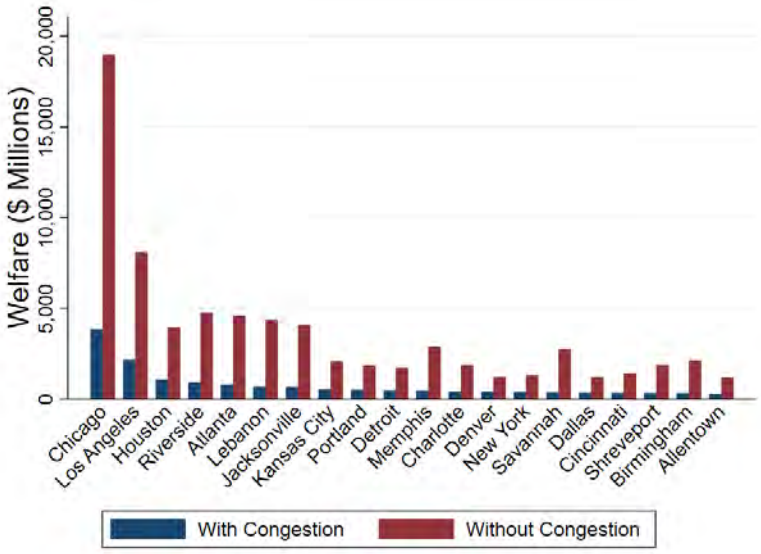
We further show that congestion plays an important role in determining the magnitudes of these welfare effects. Specifically, the benefits from improvements without congestion are 3.9 times higher at Chicago relative to improvements with congestion, 2.7 times for Los Angeles, 4.7 times for Atlanta, and 3.1 times at Kansas City (Figure 9). The weighted average benefits are 2.7 times higher without congestion compared to benefits with congestion (Figure A.6). Targeting infrastructure investments at nodes that are likely bottlenecks is crucial since these bottlenecks tend to occur due to limited capacity or a lack of alternative routes, both of which can hinder the transportation of goods. Both these factors are particularly pertinent for intermodal terminals, suggesting that alleviating congestion at terminals can yield large positive returns for the overall performance of the multimodal network.

In order to calculate the environmental impact from this counterfactual, we obtain data on transport mode-specific greenhouse gas (GHG) emissions from EPA (2022*a*). For each counterfactual improvement of the intermodal terminals, we calculate the change in mode-specific traffic flows using Equation (5.3) and multiply this change with the mode-specific GHG emissions in order to determine the change in GHG emissions.[37] We employ EPA (2022*b*)'s estimates when calculating the social cost of GHG emissions.

Table 4 reports the environmental impacts of the intermodal terminal improvements. For all terminals, this decrease in the cost of accessing multimodal transport raises the relative cost of using truck transport which decreases road traffic flows, resulting in a decrease in GHG emissions from trucks—highlighting the modal substitution effect (Table 4, Column (4)). Since the switching cost of intermodal terminals have decreased, rail traffic flows will increase in response which increases GHG emissions for rail (Table 4, Column (5)). However, since rail is much less polluting than road, the overall GHG emission change is net decrease for all the terminals. The implied welfare benefit of lowering the switching costs in some of the most central nodes could result in an unintended environmental benefit, equivalent to between $17-214 million (Table 4, Column (6)). Our findings are align with the objectives of the inter-agency U.S. National Blueprint for Transportation Decarbonatization, which prioritizes improving multimodal freight transport, and the European Green Deal, which aims to shift 75% of inland freight from road to rail and inland waterways (DOE, DOT, EPA, and HUD, 2024; European Commission, 2021).

[37] The aggregate growth from these improvements outweighs the partial equilibrium effect from mode-specific traffic changes and leads to a global increase in greenhouse gas emissions. In order to isolate the partial equilibrium effect, we keep overall (GHG weighted) traffic constant.

Figure 9. Welfare Benefits of Improving Intermodal Terminals: with and without Congestion



Notes: This figure presents a comparison of the welfare impacts for the top 20 most impactful terminals with and without congestion effects. See Figure A.6 for a scatterplot comparing both effects for all terminals.

Table. 4. Ranking: Environmental Impact of Intermodal Terminal Improvements

	(1) CBSA Name	(2) Benefit ($m)	(3) Cost ($m)	(4) Truck GHG (kt)	(5) Rail GHG (kt)	(6) GHG Benefit ($m)
1	Chicago	3851	814	78.97	151.25	214.29
2	Los Angeles	2168	537	20.42	67.98	98.21
3	Houston	1086	42	0.52	35.35	52.78
4	Riverside	942	79	25.64	38.83	54.05
5	Atlanta	811	144	58.66	30.99	38.00
6	Harrisburg	698	62	11.84	38.47	55.46
7	Jacksonville	689	38	22.83	30.71	43.01
8	Kansas City	510	81	11.26	14.97	21.01
9	Portland	492	28	6.64	14.44	20.67
10	Detroit	463	36	9.63	12.43	17.29

Notes: The table shows the environmental impact from the ten terminals where a one percent reduction of the transshipment cost generates the highest benefit. Column (1) indicates the core based statistical areas (CBSA) name of the node, which includes both metropolitan and micropolitan areas. Column (2) calculates how much 2012 U.S. GDP would need to increase in order to match the overall welfare gain (reproduced from Table 3 Column (6)), while Column (3) presents the required cost of making this one percent cost decrease (reproduced from Table 3 Column (7)). Columns (4) and (5) report changes in truck and rail GHG emissions respectively, which comes from the corresponding changes in their traffic flows and is measured in kilotons (kt). Column (6) then presents the resulting net social cost or benefit associated with these mode-specific GHG emission changes. Waterway emissions are included in Column (6) but omitted here for brevity.

6.2 Channel Decomposition Using Highway System Improvements

Our framework incorporates four main ingredients: (1) road and terminal congestion, (2) productivity and amenity spillovers, (3) multiple transport modes, and (4) endogenous routing. Guided by the theoretical discussion in Section 3.6, we decompose our results to quantify the contribution of each channel to welfare in counterfactual improvements of the U.S. highway improvement network.

Specifically, we quantify the welfare gains of reducing the transport cost by 1% on each link within the U.S. Highway System in both directions. The highway system consists of 704 highway links in total, each corresponding to a separate counterfactual. We begin by presenting our full multimodal results.

Figure 10 maps the welfare gains associated with this cost reduction across individual road links. The largest gains arise on links connecting densely populated and economically integrated regions, including corridors such as San Diego–Los Angeles, New York–Philadelphia, and major Midwestern connections such as Chicago–Rockford and Chicago–Indianapolis, as well as key routes in the southern U.S. like Durham-Chapel Hill-Raleigh. Across all links, the GDP-weighted average welfare gain, where each link is weighted by the GDP of its origin and destination nodes, is equivalent to approximately $87.7 million (in 2012 USD). Among the top 100 links, the corresponding weighted average gain rises to $137.8 million.

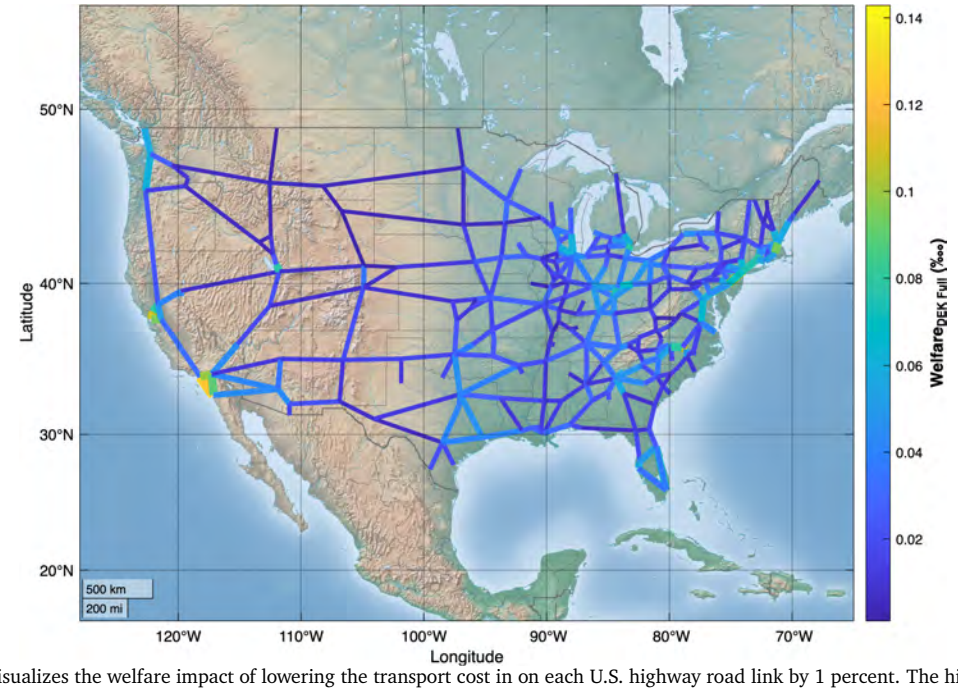
Figure 11 presents the decomposition by comparing the full multimodal framework (red line) to a specification that shuts down one of the four channels. For each specification, we run 704 counterfactuals—one for each highway link. In the scatterplot, each highway link is a dot with its welfare change plotted against the link's traffic share of GDP, which serves an indicator of its economic significance. As expected, there is a general positive relationship between a link's economic importance and its welfare gains from improvements. However, relative to the full model, shutting down individual channels alters both the slope of this relationship and the relative ranking of links. We now examine each comparison in turn.

First, shutting down road and terminal congestion in Figure 11a leads to substantially larger average welfare effects, increasing the weighted mean by 85.1%. The links with the highest welfare gains and traffic share, San Diego–Los Angeles, sees welfare increases 23-40% (depending on the direction) showing that congestion is binding for these routes. Other links in the Los Angeles area, like Los Angeles–Riverside, rise even more in rankings, with welfare gains of 62–95%, indicating that this region is particularly affected by congestion. In contrast, less-congested links experience relative declines in ranking. The difference between the full model and the no-congestion specification tends to increase with the road link's traffic share of GDP, indicating that economically important routes experience higher levels of congestion and would therefore benefit even more from infrastructure improvements if congestion were absent.[38] These quantitative findings mirror our discussion of *congestion wedges* in Section 3.6 and their tendency to attenuate the welfare effects of transportation improvements.

Second, we compare our results against a local social-savings style approximation (Hulten, 1978), where welfare gains are approximately proportional to a link's traffic and which shuts down agglomeration and dispersion externalities (productivity and amenity effects, see discussion in Section 3.6). The blue line in Figure 11b closely follows the 45-degree line through the origin, reproducing the traditional social-savings intuition that "busy roads are valuable roads." However, since our calibration implies that agglomeration and dispersion externalities are net dispersive, the relative benefit of improving road links that are economically important declines. As a result, the traditional approach overestimates average welfare effects by 57%. For example, the Washington DC–Baltimore and Chicago–Milwaukee links are ranked much higher under the traditional approach, with estimated gains exceeding 95% and 80%

[38] Shutting down terminal congestion raises estimated welfare gains by an average of 12% (Figure A.8), highlighting links connected to major intermodal terminals. We quantify large effects for links like Los Angeles-Riverside and Chicago-Milwaukee, which see welfare increases of 40.1-46.3% and 16.8-39.5% respectively.

Figure 10. Welfare Benefits of Improving Road Links



Notes: This figure visualizes the welfare impact of lowering the transport cost in on each U.S. highway road link by 1 percent. The highway system consists of 704 links in total, each corresponding to a separate counterfactual. The color scale on the right indicates the welfare gain size: the lighter the color, the larger the welfare gains. The lines indicate the graph representation of the primary road network. State boundaries are included.

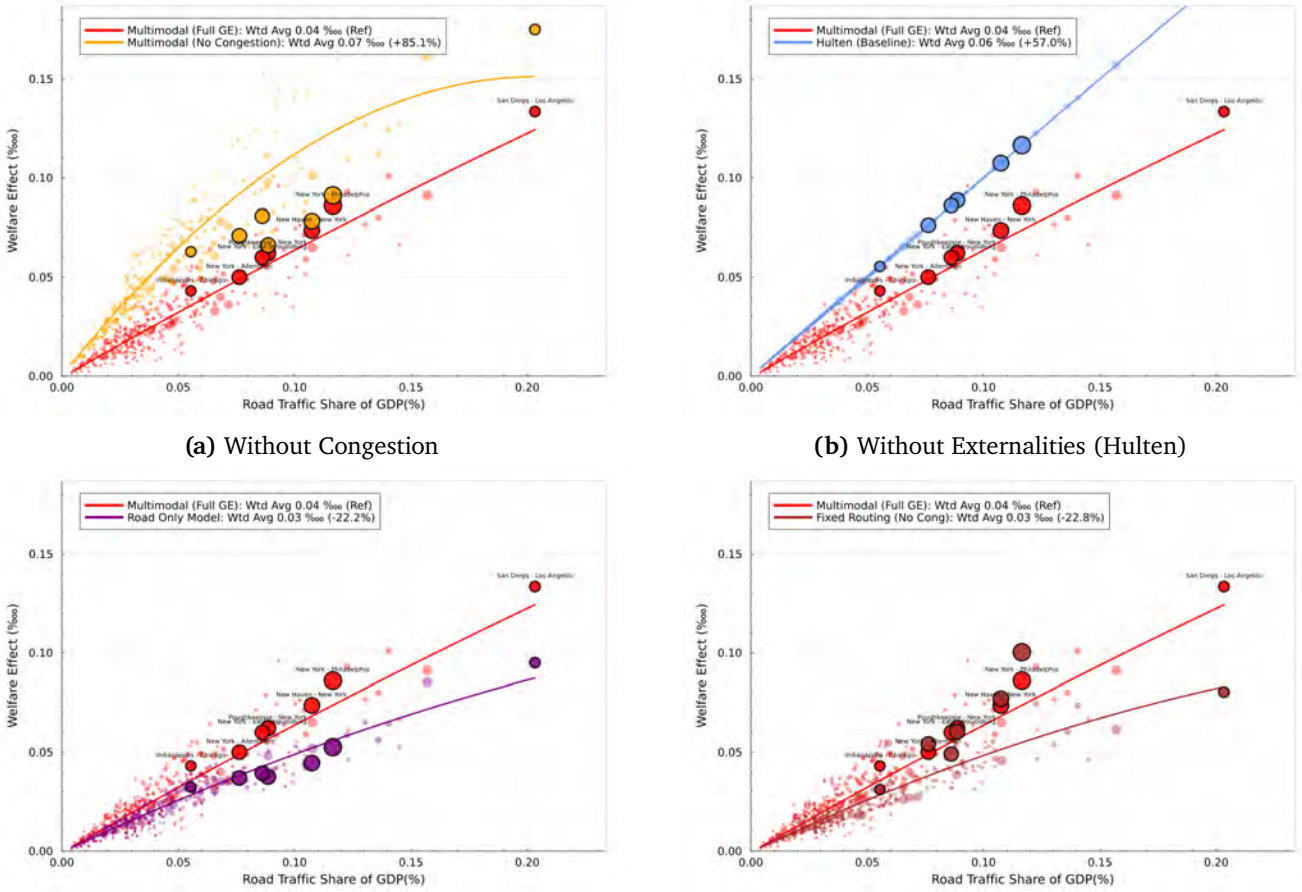
respectively relative to the full model.

Third, we examine how including multiple modes of transport changes the welfare returns to road improvements. Removing multimodal flexibility effectively collapses the network to a single mode, roads, and leads to lower average welfare effects by 22.2% compared to the full model. Road improvements can generate both modal substitution (shifting traffic from rail/barges to road) and modal complementarity (increasing overall market access via better connectivity). In our results, the complementarity channel dominates, so a unimodal framework underestimate the welfare gains because it does not take into account market access benefits and spillovers onto other modes, particularly for central links in the multimodal network. For example, welfare gains for Chicago-Indianapolis, San Diego-Los Angeles, and New York-Philadelphia are 58.9%, 43.1%, and 74% lower, respectively, in the unimodal scenario compared with the full model. This is consistent with our previous results from Section 6.1 showing that improving the intermodal terminals in these locations generate the highest welfare impacts (Figure 8). Additionally, these quantitative findings reinforce our discussion of mode *reallocation wedges* in Section 3.6: allowing multiple modes creates *net modal complementarity* which reinforce the value of the improved road links through expanded access and spillovers.

Finally, we implement a version of the model that determines welfare gains, but keeping route structure as fixed thus determining the additional benefit from modeling the additional flexible adjustment of traffic as infrastructure improves.[39] Eliminating endogenous routing changes the spatial distribution

[39] See Appendix A.13 for detailed derivations.

Figure 11. Decomposition of Main Channels: Welfare Benefits of Improving Road Links



(a) Without Congestion

(b) Without Externalities (Hulten)

(c) Without Multimodal Flexibility

(d) Without Endogenous Routing

Notes: This figure reports the welfare effects of a 1% reduction in transport costs on each link of the U.S. highway system and decomposes the main mechanisms in the model. The highway system consists of 704 links in total, each corresponding to a separate counterfactual. Each model specification evaluates the same set of 704 link-level counterfactuals. The red line in each panel represents the full multimodal model, which includes road and terminal congestion, productivity and amenity spillovers (Hulten effects), multiple transport modes, and endogenous routing. Each panel compares the full model to a specification that shuts down one channel: congestion (Figure 11a), externalities (Figure 11b), multimodal flexibility (unimodal road-only network, Figure 11c), or endogenous routing (Figure 11d). The legend in each panel reports the weighted average welfare effect for each specification, with weights given by the GDP of origin and destination nodes of each link). The percentage in parentheses indicates the difference relative to the full multimodal model. For example, abstracting from congestion increases the average estimated welfare gain by 85.1% relative to the full model (Figure 11a). Figure A.7 presents all decomposition results jointly in a single figure.

of benefits between links that are central to the routing network and those that are less central (Figure 11d). Links that are central to the network, meaning ones used by many other routes, experience larger welfare gains under the full model because traffic can increase in response to improvements on these links. Under fixed (exogenous) routing, these links cannot attract additional traffic, so their estimated welfare gains are substantially smaller, 35–43% lower for San Diego–Los Angeles and roughly 35–37% lower for Chicago–Rockford and Chicago–Indianapolis. Conversely, links connecting urban centers such as New York–Philadelphia experience even larger welfare gains under fixed routing since traffic cannot adjust. Overall, fixing routes underestimates the average welfare effects by 22.8% relative to the full

model. This underestimate is similar in magnitude to the underestimate from ignoring multimodal flexibility, suggesting that most rerouting responses to highway improvements take place across modes rather than within modes. These results are aligned with our discussion of route *reallocation wedges* in Section 3.6, showing that endogenous rerouting amplifies welfare gains for centrally positioned links.

Our estimates are complementary to and build upon existing literature. AA2022 shows that ignoring road traffic congestion overstates welfare gains from infrastructure improvements, particularly on busy segments. We find that this is the case in our setting as well, with the additional compounding role of terminal congestion. Fan, Lu and Luo (2019) documents that ignoring rerouting and the substitution between expressways and regular roads in China can underestimate welfare loss from the removal of expressways by 21%-46%. In the U.S. context and and across three transport modes (highways, rail, and barges), we separately quantify these two channels and find similarly large biases (underestimate of 22.2% without modal substitution and underestimate of 22.5% without endogenous routing).[40] Consistent with this, Foellmi, Hepenstrick and Torun (2024) also shows that standard approaches with fixed routing underestimate welfare gains from trade cost reductions relative to models that allow for rerouting consistent with the triangle inequality.[41] Using a sufficient statistics approach, Allen, Fuchs and Wong (2025) shows that traditional approaches overestimate welfare gains from road improvements without accounting for externalities and traffic congestion. This is consistent with our finding where we separately quantify the overestimate for each of these channels (overestimate of 85.1% without congestion and overestimate of 57% without externalities).

6.3 Value of the Multimodal System: Impact from Losing Rail Network Access

In the fall of 2022, a nationwide rail strike that would have shut down the U.S. rail system was narrowly averted following intervention by U.S. President Biden (Kanno-Youngs and Cochrane, 2022). In order to evaluate the impact of losing access to the multimodal rail network, we estimate the aggregate welfare impact of increasing the cost on all rail links to prohibitively high levels (twenty-fold). The first row in Table 5 shows that the welfare cost of losing the multimodal rail network is equivalent to decreasing U.S. GDP by about $230 billion (in 2012 USD). While existing literature have quantified the importance of the road and highway transport network, the value of the rail network has received less attention except in historical analyses. Our estimate of the value of the contemporary rail network for container freight is roughly half the value of the U.S. highway system as estimated by Jaworski, Kitchens and Nigai (2023), and about 1.8 times the historic inflation-adjusted value of railroads to agricultural land value

[40] In a historical context, Donaldson and Hornbeck (2016) estimates that removing U.S. railroads would reduce agricultural land values in 1890 by 60.2%. Modal substitution through an expanded canal network and improved country roads would offset roughly 34% of these losses. Our modal substitution decomposition results are broadly consistent with these findings, although our focus is specifically on containerized freight.

[41] Foellmi, Hepenstrick and Torun (2024) develops a trade cost estimation procedure across countries that allows for rerouting which satisfies the triangle inequality.

as estimated by Donaldson and Hornbeck (2016).

How much does congestion affect these results? We next estimate the same increase in rail link costs, but reducing the congestion at intermodal terminals to zero ($\lambda_m = 0$). We find a very small increase in the negative welfare impact of 2.6%, which is expected since the terminal congestion effects are not binding due to prohibitively high cost of rail—there is very little traffic flows on the rail network. The small increase is likely due to intermodal terminals that are also ports. However, once all congestion effects are removed, the negative impact grows by an additional 18%, equivalent to decreasing U.S. GDP by about $279 billion in 2012 USD, first row of Table 5, Column (4)). Consistent with our terminal improvement counterfactuals, congestion has a compounding effect on welfare (Figures 9 and A.6).

Next, we evaluate the environmental consequences of the rail strike. Under this scenario, cargo that would typically move by rail is diverted onto the road network, leading to a large increase in truck GHG emissions of more than 38,000 kt (first row of Table 6, Column (2)). Although rail GHG emissions decline due to the absence of rail transport (first row, Column (3)), trucks are far more polluting per ton-mile. The resulting net social cost from these emission changes is $11.9 billion, on top of the initial welfare loss of $230 billion (first row of Table 6, Columns (4) and (5)). Overall, the negative welfare impact of losing access to the multimodal rail network is compounded by the additional environmental damage from the modal substitution towards truck transport.

6.4 Removal of Jones Act: U.S.-Owned, -Built, and -Crewed Shipping Requirement

Next, we quantify the consequences of repealing the Jones Act, the Merchant Marine Act of 1920. This is an active U.S. trade regulation which requires all ships transporting freight and passengers between U.S. ports to be built in the U.S., owned by U.S. citizens, and mostly crewed by U.S. citizen. Jones Act ships on average cost 2.7 times to operate than foreign-flag equivalents (MARAD, 2011).[42] We evaluate the repeal of the Jones Act by decreasing the link cost of waterways by the MARAD (2011) estimate. The second row in Table 5 shows that the welfare gain of repealing the Jones Act and allowing foreign ships to transport cargo between U.S. ports is equivalent to increasing U.S. GDP by about $3.2 billion (in 2012 USD). Our estimate lies towards the lower end of the range of previous federal government estimates: $2.8 billion (USITC, 1995) to $9.8 billion (USITC, 1991). One reason is scope: we evaluate only the continental United States. The Jones Act repeal will have much larger benefits for Hawaii, Alaska, and Puerto Rico since these locations rely a lot more on maritime transport and have limited substitutes to other modes. Another potential reason is that our estimates reflect the U.S. multimodal network today, which has had more than a century—since the Jones Act's passage in 1920—to adjust away from costly coastal shipping. Dense rail and road networks run up and down the U.S. coastlines in present day

[42] Olney (2020) studies the economic costs of the Jones Act and shows that U.S. container ships cost 5 times more to build than foreign ships. Using United Kingdom shipbuilding as an instrument, this paper finds that the Jones Act disproportionately restricts domestic U.S. water trade especially in coastal states.

Table. 5. Welfare Impact of Policy Scenarios: The Role of Congestion

(1) Scenario	(2) Full Welfare ($bn)	(3) No Terminal Cong. ($bn)	(4) No Cong. ($bn)
Value of the Multimodal System	-230.46	-236.40	-278.94
Removal of the Jones Act	3.15	11.73	16.16
Panama Canal Disruptions	-2.67	-7.64	-10.29

Notes: The table reports the welfare impact of three policy scenarios. The first row quantifies the loss of rail network access, motivated by the averted U.S. rail strike in fall 2022. The second row quantifies the repeal of the Merchant Marine Act of 1920 (Jones Act), which restricts domestic maritime shipping to vessels that are U.S.-owned, -built, and -crewed. The third row quantifies disruptions to the Panama Canal due to drought conditions. Column (1) labels each scenario. Column (2) reports how much 2012 U.S. GDP would need to change in order to match the welfare change in the scenario. Column (3) reports the same calculation assuming no congestion at intermodal terminals while Column (4) reports the same calculation assuming no congestion at both intermodal terminals and roads.

Table. 6. Environmental Impact of Policy Scenarios

	(1) Scenario	(2) Truck GHG Change (kt)	(3) Rail GHG Change (kt)	(4) GHG Benefit ($bn)	(5) Benefit ($bn)
1	Railroad Strike	38947	-5171	-11.88	-230.46
2	Removal of the Jones Act	-589	-47	0.19	3.15
3	Panama Canal	1524	111	-0.45	-2.67

Notes: The table reports the environmental impact of three policy scenarios. The first row quantifies the loss of rail network access, motivated by the averted U.S. rail strike in fall 2022. The second row quantifies the repeal of the Merchant Marine Act of 1920 (Jones Act), which restricts domestic maritime shipping to vessels that are U.S.-owned, -built, and -crewed. The third row quantifies disruptions to the Panama Canal due to drought conditions. Column (1) labels each scenario. Columns (2) and (3) report changes in truck and rail GHG emissions respectively, which comes from the corresponding changes in their traffic flows and is measured in kilotons (kt). Column (4) then presents the resulting net social cost or benefit associated with the mode-specific GHG emission changes. Column (5) calculates how much 2012 U.S. GDP would need to change in order to match the overall welfare change from each scenario (reproduced from Table 5 Column (2)). Waterway emission changes are included in Column (4) but omitted here for brevity.

(Figure 3). Had waterborne transport been less expensive, these networks might well have been built further inland instead.

In this scenario, the congestion effects of intermodal terminals play a much larger role in informing the welfare estimates. Without intermodal terminal congestion, the welfare benefits from repealing the Jones Act increase by 272% (second row in Table 5 Column (3)). This is due to the gains from being able to increase waterway transport use once it is cheaper. Without all congestion (road and terminals), the welfare gains from the repeal increase by more than 4 times (second row in Table 5 Column (4)).

Additionally, the repeal of the Jones Act decreases emissions overall. This is because cargo would substitute towards using waterways due to the lower cost, and so both road and rail traffic would decrease as a result (second row of Table 6, Columns (2) and (3)). Since barge emissions are much lower, this results in a GHG benefit increase of 0.2 billion dollars (second row of Table 6, Column (4)).

6.5 Panama Canal Disruptions: Impact of Drought Conditions

Our third scenario studies the impact of the drought conditions in the Panama Canal, which decreases the number of ships that can access it (Rojanasakul, 2024). To trace out the importance of the Panama Canal and how it interacts with the U.S. multimodal transportation system, we include in our calibration a (coarse) notion of routed imports at the port-level. Specifically, we include linkages from ports to foreign destinations and calibrate the importance of these linkages with port-destination specific trade

flows.[43] To simulate the impact of the Panama Canal, we assume that routes that typically require usage of the Panama Canal (e.g. Asia to ports on the East Coast) are affected by transport cost increase. We evaluate this disruption by increasing the cost of accessing the canal by five times.

The third row in Table 5 shows that the welfare cost of disruptions at the Panama Canal is equivalent to decreasing U.S. GDP by about $2.7 billion (in 2012 USD). To the best of our knowledge, this is the first estimate of this disruption allowing for modal and route substitution, including at ports. Here, removing congestion again compounds these effects. Without terminal congestion, the effects increase by more than 186%. This large increase is driven by the modal substitution from water to rail and road, and the intermodal terminals playing a crucial role in facilitating this transfer. Without terminal and road congestion, the increase is a further 35 percent. The modal substitution due to the canal disruptions in turn increases the use of both truck and rail which increases GHG emissions on both modes (third row of Table 6). This results in an overall decrease in the GHG benefit of 0.5 billion dollars (third row of Table 6, Column (4)).

7 Conclusion

Freight transportation is fundamentally multimodal, yet most quantitative analyses of transport infrastructure still abstract from either transport mode choice, intermodal bottlenecks, or both. In this paper, we develop a tractable spatial equilibrium framework that embeds recursive route choice, nested mode choice, and mode-specific congestion into a unified model of the freight network. We then bring this framework to the data by estimating an elasticity of modal substitution between rail and truck and a congestion elasticity at intermodal terminals using vessel-positioning data. This combination of theory, estimation, and calibration allows us to evaluate infrastructure improvements and disruptions in a way that is both economically disciplined and operationally realistic.

Our results show that the multimodal structure of the transport system matters quantitatively for both welfare and emissions. Improvements at intermodal terminals generate large gains, especially at centrally located hubs such as Chicago, Atlanta, and Kansas City, because these nodes connect coastal gateways to the interior and mediate substitution across road, rail, and waterways. At the same time, the paper shows that congestion substantially attenuates the gains from infrastructure improvements: abstracting from congestion leads to sizable overstatements of the welfare effects of highway investments, while abstracting from multimodal flexibility or endogenous rerouting leads to understatements. More broadly, the decomposition makes clear that the value of a local improvement depends not only on the traffic it directly carries, but also on how that improvement reorganizes traffic across the rest of the network and on whether those induced adjustments run into scarcity at roads or terminals.

[43] We collect data of destination specific imports from the Census Bureau and aggregate across five major regions: Central America, Africa, Asia, Australia and Oceania, Europe, North America (excl U.S.A), South/Central America.

Our policy counterfactuals highlight the same message from a different angle. The multimodal rail network is valuable: shutting it down generates welfare losses on the order of $ 230 billion in GDP-equivalent terms, with additional environmental costs as freight shifts toward road transport. Conversely, reducing barriers to waterborne domestic transport through repeal of the Jones Act raises welfare, while the Panama Canal disruptions impose meaningful losses that are amplified by congestion at the terminal nodes needed to absorb diverted traffic. Taken together, these results imply that the economic cost of disruptions depend on the extent of substitution into other modes, the capacity of intermodal terminals to accommodate that substitution, and the broader equilibrium effects on market access across space.

The broader implication is that intermodal terminals should be viewed as core infrastructure in their own right. They are not merely transfer points between networks. Instead, they are locations where the efficiency, resilience, and environmental performance of the freight system are jointly determined. This makes them natural targets for transport infrastructure policy, especially in an environment where geopolitical risk, climate-related disruptions, and decarbonization goals are becoming increasingly salient. By showing how multimodal substitution, congestion, and the network structure interact, our framework provides a way to rank such investments and to quantify the trade-offs between efficiency, robustness, and emissions reduction. Future work could extend this framework to incorporate additional margins, such as inventories, market power, scheduling and reliability, or international supply-chain propagation. The key insight remains: the welfare effects of transport infrastructure investments depend on the multimodal network through which goods actually move.

References

Akbar, Prottoy A, Victor Couture, Gilles Duranton, and Adam Storeygard. 2023. "The fast, the slow, and the congested: Urban transportation in rich and poor countries." National Bureau of Economic Research.

Albouy, David, Gabriel Ehrlich, and Minchul Shin. 2018. "Metropolitan land values." *Review of Economics and Statistics*, 100(3): 454–466.

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

Allen, Treb, and Costas Arkolakis. 2022. "The Welfare Effects of Transportation Infrastructure Improvements." *The Review of Economic Studies*.

Allen, Treb, and Costas Arkolakis. 2025. "Quantitative regional economics." National Bureau of Economic Research Working Paper.

Allen, Treb, Simon Fuchs, and Woan Foong Wong. 2025. "Evaluating Transportation Improvements Quantitatively: A Primer."

Almagro, Milena, Felipe Barbieri, Juan Camilo Castillo, Nathaniel G Hickok, and Tobias Salz. 2024. "Optimal urban transportation policy: Evidence from Chicago." National Bureau of Economic Research.

Anderson, James E., and Eric van Wincoop. 2003. "Gravity with Gravitas: A Solution to the Border Puzzle." *American Economic Review*, 93(1): 170–192.

Antràs, Pol, and Alonso Gortari. 2020. "On the geography of global value chains." *Econometrica*, 88(4): 1553–1598.

Asturias, Jose. 2020. "Endogenous transportation costs." *European economic review*, 123: 103366.

Baum-Snow, Nathaniel. 2007. "Did highways cause suburbanization?" *The quarterly journal of economics*, 122(2): 775–805.

Beuthe, Michel, Bart Jourquin, and Natalie Urbain. 2014. "Estimating freight transport price elasticity in multi-mode studies: A review and additional results from a multimodal network model." *Transport Reviews*, 34(5): 626–644.

Bonadio, Barthélémy. 2021. "Ports vs. Roads: Infrastructure, Market Access and Regional Outcomes." Working Paper.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2025. "A practical guide to shift-share instruments." *Journal of Economic Perspectives*, 39(1): 181–204.

Boyles, Stephen D., Nicholas E. Lownes, and Avinash Unnikrishnan. 2021. *Transportation Network Analysis*. Vol. 1. 0.89 ed.

Brancaccio, Giulia, Myrto Kalouptsidi, and Theodore Papageorgiou. 2020. "Geography, transportation, and endogenous trade costs." *Econometrica*, 88(2): 657–691.

Brancaccio, Giulia, Myrto Kalouptsidi, and Theodore Papageorgiou. 2024. "Investment in Infrastructure and Trade: The Case of Ports." National Bureau of Economic Research Working Paper.

Brooks, Leah, Nicolas Gendron-Carrier, and Gisela Rua. 2018. "The local impact of containerization." *Finance and Economics Discussion Series*, 45.

Carballo, Jeronimo, Alejandro Graziano, Georg Schaur, and Christian Volpe Martincus. 2021. "Import Processing and Trade Costs." CESifo Working Paper no. 9170.

CBO. 2022. "Emissions of Carbon Dioxide in the Transportation Sector." Congressional Budget Office Publication 58566.

Census Bureau. 2003-2021. "USA Trade Online." *https://usatrade.census.gov*, Last accessed on August 2024.

Coşar, A Kerem, and Banu Demir. 2018. "Shipping inside the box: Containerization and trade." *Journal of International Economics*, 114: 331–345.

Couture, Victor, Gilles Duranton, and Matthew A Turner. 2018. "Speed." *Review of Economics and Statistics*, 100(4): 725–739.

Cristea, Anca, David Hummels, Laura Puzzello, and Misak Avetisyan. 2013. "Trade and the greenhouse gas emissions from international freight transport." *Journal of environmental economics and management*, 65(1): 153–173.

Daly, Andrew, and Michel Bierlaire. 2006. "A general and operational representation of Generalised Extreme Value models." *Trans. Res. Part B: Methodol.*, 40(4): 285–305.

Degiovanni, Pedro, and Ron Yang. 2023. "Economies of Scale and Scope in Railroading." Working Paper.

Dekle, Robert, Jonathan Eaton, and Samuel Kortum. 2008. "Global Rebalancing with Gravity: Measuring the Burden of Adjustment." National Bureau of Economic Research, Inc NBER Working Papers 13846.

Department of Transportation. 2017-2024. "Value, Tonnage, and Ton-Miles of Freight by Distance Band." https://www.bts.gov/browse-statistical-products-and-data/freight-facts-and-figures/value-tonnage-and-ton-miles-freight, Freight Facts and Figures, Bureau of Transportation Statistics, U.S. Department of Transportation.

Djankov, Simeon, Caroline Freund, and Cong S Pham. 2010. "Trading on time." *The review of Economics and Statistics*, 92(1): 166–173.

Do, Anh D, Sharat Ganapati, Woan Foong Wong, and Oren Ziv. 2025. "Transshipment hubs, trade, and supply chains." National Bureau of Economic Research.

DOE, DOT, EPA, and HUD. 2024. "US National Blueprint for Transportation Decarbonatization." *www.energy.gov/sites/default/files/2023-01/the-us-national-blueprint-for-transportation-decarbonization.pdf*, US Department of Energy, US Department of Transportation, US Environmental Protection Agency, US Department of Housing and Urban Development.

Donald, Eric, Masao Fukui, and Yuhei Miyauchi. forthcoming. "Unpacking Aggregate Welfare in a Spatial Economy." *Review of Economic Studies*. Forthcoming.

Donaldson, Dave. 2025. "Transport infrastructure and policy evaluation." In *Handbook of Regional and Urban Economics*. Vol. 6, 287–352. Elsevier.

Donaldson, Dave, and Richard Hornbeck. 2016. "Railroads and American economic growth: A "market access" approach." *The Quarterly Journal of Economics*, 131(2): 799–858.

Ducruet, César, Réka Juhász, Dávid Krisztián Nagy, and Claudia Steinwender. 2024. "All aboard: The effects of port development." *Journal of International Economics*, 151: 103963.

Dunn, Jason, and Fernando Leibovici. 2023. "Navigating the Waves of Global Shipping: Drivers and Aggregate Implications." Federal Reserve Bank of St. Louis, Research Division.

Duranton, Gilles, and Diego Puga. 2023. "Urban growth and its aggregate implications." *Econometrica*, 91(6): 2219–2259.

Duranton, Gilles, and Matthew A Turner. 2011. "The fundamental law of road congestion: Evidence from US cities." *American Economic Review*, 101(6): 2616–52.

Duranton, Gilles, and Matthew A Turner. 2012. "Urban growth and transportation." *Review of Economic Studies*, 79(4): 1407–1440.

Duranton, Gilles, Peter M Morrow, and Matthew A Turner. 2014. "Roads and Trade: Evidence from the US." *Review of Economic Studies*, 81(2): 681–724.

Eaton, Jonathan, and Samuel Kortum. 2002. "Technology, Geography, and Trade." *Econometrica*, 70(5): 1741–1779.

Economides, Philip. 2024. "Unconventional Protectionism in Containerized Shipping." *Available at SSRN 4930357*.

EPA. 2022*a*. "Inventory of US Greenhouse Gas Emissions and Sinks: 1990-2020." *https://www.epa.gov/ghgemissions/draft-inventory-us-greenhouse-gas-emissions-and-sinks-1990-2020*, US Environmental Protection Agency, EPA 430-R-22-0003.

EPA. 2022*b*. "Report on the social cost of greenhouse gases: Estimates incorporating recent scientific advances." US Environmental Protection Agency Washington, DC.

EPA. 2024. "Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2022." U.S. Environmental Protection Agency, EPA 430-R-24-004.

European Commission. 2021. "Sustainable and Smart Mobility Strategy." *https://transport.ec.europa.eu/system/files/2021-04/2021-mobility-strategy-and-action-plan.pdf*, Directorate-General for Mobility and Transport.

Fajgelbaum, Pablo D, and Edouard Schaal. 2020. "Optimal transport networks in spatial equilibrium." *Econometrica*, 88(4): 1411–1452.

Fajgelbaum, Pablo D, Cecile Gaubert, Nicole Gorton, Eduardo Morales, and Edouard Schaal. 2023. "Political Preferences and the Spatial Distribution of Infrastructure: Evidence from California's High-Speed Rail." National Bureau of Economic Research.

Fan, Jingting, and Wenlan Luo. 2020. "A Tractable Model of Transshipment." Working Paper.

Fan, Jingting, Yi Lu, and Wenlan Luo. 2019. "Valuing domestic transport infrastructure: a view from the route choice of exporters." *The Review of Economics and Statistics*, 1–46.

Foellmi, Reto, Christian Hepenstrick, and David Torun. 2024. "Triangle inequalities in international trade: The neglected dimension." *Journal of International Economics*, 152: 104018.

Ganapati, Sharat, Woan Foong Wong, and Oren Ziv. 2021. "Entrepot: Hubs, scale, and trade costs." National Bureau of Economic Research Working Paper 29015.

Gibbons, Stephen, Stephan Heblich, and Edward W Pinchbeck. 2024. "The spatial impacts of a massive rail disinvestment program: The Beeching Axe." *J. Urban Econ.*, 143(103691): 103691.

Harrigan, James. 2010. "Airplanes and comparative advantage." *Journal of International Economics*, 82(2): 181–194.

Head, Keith, and Thierry Mayer. 2014. "Gravity equations: Workhorse, toolkit, and cookbook." In *Handbook of international economics*. Vol. 4, 131–195. Elsevier.

Heiland, Inga, Andreas Moxnes, Karen-Helene Ulltveit-Moe, and Yuan Zi. 2019. "Trade From Space: Shipping Networks and The Global Implications of Local Shocks." C.E.P.R. Discussion Papers CEPR Discussion Papers 14193.

Horn, Roger A, and Charles R Johnson. 2012. *Matrix analysis*. Cambridge university press.

Hulten, Charles R. 1978. "Growth accounting with intermediate inputs." *The Review of Economic Studies*, 45(3): 511–518.

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

Jaworski, Taylor, Carl Kitchens, and Sergey Nigai. 2023. "Highways and globalization." *International Economic Review*, 64(4): 1615–1648.

Kanno-Youngs, Zolan, and Emily Cochrane. 2022. "Biden Signs Legislation to Avert Nationwide Rail Strike." *https://www.nytimes.com/2022/12/02/us/politics/rail-strike-biden.html*, New York Times.

Lind, Nelson, and Natalia Ramondo. 2023. "Trade with Correlation." *American Econonomic Review*, 113(2): 317–353.

Ludwig, Philipp. 2025. "Can Unilateral Policy Decarbonize Maritime Trade?" CESifo Working Paper.

Lugovskyy, Volodymyr, Alexandre Skiba, and David Terner. 2022. "Unintended Consequences of Environmental Regulation of Maritime Shipping: Carbon Leakage to Air Shipping." Working Paper.

MARAD. 2011. "Comparison of US and Foreign-flag Operating Costs." US Department of Transportation, Maritime Administration.

McFadden, Daniel, Clifford Winston, and Axel Boersch-Supan. 1986. "Joint estimation of freight transportation decisions under nonrandom sampling." *Analytical Studies in Transport Economics*, , ed. Andrew F.Editor Daughety, 137–158. Cambridge University Press.

Melo, Emerson. 2012. "A representative consumer theorem for discrete choice models in networked markets." *Econ. Lett.*, 117(3): 862–865.

Michaels, Guy. 2008. "The effect of trade on the demand for skill: Evidence from the interstate highway system." *The Review of Economics and Statistics*, 90(4): 683–701.

Mundaca, Gabriela, Jon Strand, and Ian R Young. 2021. "Carbon pricing of international transport fuels: Impacts on carbon emissions and trade activity." *Journal of Environmental Economics and Management*, 110: 102517.

National Economic Council. 2024. "2021–2024 Quadrennial Supply Chain Review."

Notteboom, Theo. 2018. "Trade and transport modes: the search for global connectivity through transport networks." In *Handbook of international trade and transportation*. 541–565. Edward Elgar Publishing.

Office of the State Auditor. 2019. "Capitalization and Depreciation of Infrastructure." Society for Economic Dynamics 2019 Meeting Papers 212.

Olney, William W. 2020. "Cabotage sabotage? The curious case of the Jones Act." *Journal of International*

Economics, 127: 103378.

Oum, Tae Hoon. 1979. "A cross sectional study of freight transport demand and rail-truck competition in Canada." *The Bell Journal of Economics*, 463–482.

Oum, Tae Hoon. 1989. "Alternative demand models and their elasticity estimates." *Journal of Transport Economics and policy*, 163–187.

Oyama, Yuki, Yusuke Hara, and Takashi Akamatsu. 2022. "Markovian traffic equilibrium assignment based on network generalized extreme value model." *Trans. Res. Part B: Methodol.*, 135–159.

Papola, Andrea, and Vittorio Marzano. 2013. "A Network Generalized Extreme Value Model for Route Choice Allowing Implicit Route Enumeration." *Comput.-aided civ. infrastruct. eng.*, 28(8): 560–580.

Ranieri, Luigi, Salvatore Digiesi, Bartolomeo Silvestri, and Michele Roccotelli. 2018. "A review of last mile logistics innovations in an externalities cost reduction vision." *Sustainability*, 10(3): 782.

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

Rich, J., O. Kveiborg, and C.O. Hansen. 2011. "On structural inelasticity of modal substitution in freight transport." *Journal of Transport Geography*, 19(1): 134–146.

Rodrigue, Jean-Paul. 2020. *The geography of transport systems.* Routledge.

Rojanasakul, Mira. 2024. "Panama Canal Drought Slows Cargo Traffic."

Santamaria, Marta. 2020. "Reshaping Infrastructure: Evidence from the division of Germany." *The Warwick Economics Research Paper Series (TWERPS)*.

Schmidt, Gregory. 2021. "A record number of cargo ships off the California coast shows a crack in the supply chain." *https://www.nytimes.com/2021/09/23/business/cargo-ships-supply-chain.html*, New York Times.

Shapiro, Joseph S. 2016. "Trade Costs, CO2, and the Environment." *American Economic Journal: Economic Policy*, 8(4): 220–54.

Staiger, Douglas, and James H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica*, 65(3): 557–586.

Surface Transportation Board. 2009. "A Study of Competition in the US Freight Railroad Industry and Analysis of Proposals that might Enhance Competition." Report prepared by Laurits R. Christensen Associates.

Tolva, Edoardo. 2025. "One way or another: Modes of transport and international trade." Technical report.

United Nations. 2021. "Climate Change Fact Sheet." United Nations Sustainable Transport Conference.

USITC. 1991. "The Economic Effects of Significant US Import Restraints." *Investigation No.332-325, Publication 3201*.

USITC. 1995. "The Economic Effects of Significant US Import Restraints."

Winston, Clifford. 1981. "A Disaggregate Model of the Demand for Intercity Freight Transportation." *Econometrica*, 49(4): 981–1006.

World Bank. 2023. "Connecting to compete 2023: trade logistics in an uncertain global economy."

Zgonc, Borut, Metka Tekavčič, and Marko Jakšič. 2019. "The impact of distance on mode choice in freight transport." *European Transport Research Review*, 11(1): 1–18.

Part

Online Appendix

Table of Contents

A Theoretical Derivations **A2**

A.1 Derivations for Section 3.2: Recursive Routing Choice A2

A.2 Derivations for Section 3.3: Combined Routing and Sourcing Choice, and Gravity A4

A.3 Derivations for Section 3.3: Combined Routing and Sourcing Choice, and Recursive Equilibrium A7

A.4 Derivations for Section 3.4: Nested Mode Choice and Aggregate Transport Cost A9

A.5 Isomorphism with Allen and Arkolakis (2022) A10

A.6 Derivations for Section 3.5: Congestion and Traffic Flows A12

A.7 Derivations for Section 3.6: Traffic Substitution and Complementarity A18

A.8 Derivations for Section 3.6: Welfare Elasticities, Wedges, and Signs A19

A.9 Derivations for Section 5: Implied Modal Wedge A27

A.10 Derivations for Section 5.1: Proofs of Proposition 1 . . . . . A28

A.11 Derivations for Section 5.3: Predicted vs Observed Transport Mode-level Trade Flows . . A33

A.12 Derivations for Section 5.1: International Trade A36

A.13 Derivations for Section 6.2: Counterfactuals with Fixed-Routing A41

B Data and Additional Results **A44**

B.1 Data Appendix A44

B.2 Regression design for Modal Substitution (Subsection 4.1) A46

B.3 Elasticity of Modal Substitution: Additional Results A49

B.4 Estimation of Intermodal Terminal Congestion: Additional Results A53

B.5 Rail Intermodal Terminal Congestion and Port Traffic A54

B.6 Model Fit: Additional Results A56

B.7 Gains from Improving Intermodal Terminals Counterfactuals: Additional Results A57

B.8 Channel Decomposition Using Highway System Improvements: Additional Results A59

A Theoretical Derivations

This appendix presents derivations for the theory section. Table A.1 lists the notations that are used in the theory framework.

Table. A.1. Overview of notation

Object	Notation
Multi-layered graph	$\mathscr{G} \equiv (\mathscr{N}, \mathscr{L})$
Successor nodes	$\mathscr{B}(i)$
Predecessor nodes	$\mathscr{F}(i)$
Neighboring nodes	$\mathscr{N}(i)$
Route of length K	$r \equiv \{i = r_0, r_1, \ldots, r_K = j\}$
Elasticity of substitution across goods	σ
Route dispersion parameter	θ
Mode dispersion parameter	η
Expected minimal transport cost between i and j	τ_{ij}
Expected minimal edge-specific transport cost between i and k	t_{ik}
Mode-specific transport cost between i and k for mode m	$\tilde{t}_{ik,m}$
Link intensity	π_{ij}^{kl}
Link-sourcing probability	$\pi_{ij,k}$
Link choice probability conditional on sourcing from i to j	π_{ij}^{k}
Mode choice probability conditional on link ik	$\pi_{ij,k}^{m}$
Mode-link-sourcing probability	$\pi_{ij,k,m}$

A.1 Derivations for Section 3.2: Recursive Routing Choice

A consumer resides in location j and makes a route-sourcing choice comparing prices across all producer locations i and transportation costs across multiple routes r. We make the assumption that transport cost are multiplicative and that they are furthermore the product of edge-specific transport costsalong the route r, i.e. $\tau_{ij,r}^{-\theta_j} = \left(\prod_{l=1}^{K} t_{r_{t-1},r_l}^{-\theta_j}\right)$. Finally, we assume that the consumer makes a recursive route-sourcing choice, beginning at the destination location j by comparing the sourcing prices across the set of (predecessor) neighboring nodes $k \in \mathscr{B}(j)$ net of the transport cost of traversing via the set of (predecessor) neighboring nodes, t_{kj}. For now - without loss of generality - we simply assume that the consumer faces a routing problem along a single-layered graph. As such, the edge-specific transport cost t_{kj} can be interpreted as the expected minimum link cost of traversing to the neighboring node along a multimodal transport network. In summary, the consumer at location j faces the following set of (recursively defined) prices:

$$p_{ij,k}(\nu) = \frac{t_{kj}\tau_{ik}w_i}{\varepsilon_{ij,k}(\nu)}$$

where $\varepsilon_{ij,k}(\nu)$ is a random variable drawn from a Frechet distribution with cumulative distribution given by,

$$F_{ijk}(\epsilon) = e^{-T_{ijk}\epsilon^{-\theta_j}}$$

The consumer in location j is presented with a set of route-source specific prices across producer locations i and chooses

a route by traversing towards neighboring node k, i.e.

$$\begin{aligned} G_{ijk}(p) &= \Pr\left(P_{ijk} \leq p\right) = 1 - F_{ijk}\left(\frac{t_{kj}\tau_{ik}w_i}{p}\right) \\ &= 1 - e^{-T_{ijk}\left(t_{kj}\tau_{ik}w_i\right)^{-\theta_j}p^{\theta_j}} \end{aligned}$$

We can use this distribution to characterize the lowest route-source specific price that consumer in location j faces. To do so, fix an arbitrary threshold price p. The lowest price will be less than p, unless each route-source specific price is greater than p. We therefore seek to characterize, $G_{ijk}(p) = \Pr\left(P_{ijk} \leq p\right)$, which is given by,

$$\begin{aligned} G_j(p) &= 1 - \prod_{i,k}\left(1 - G_{ijk}(p)\right) \\ &= 1 - \prod_{i,k} e^{-T_{ijk}\left(t_{kj}\tau_{ik}w_i\right)^{-\theta_j}p^{\theta_j}} \\ &= 1 - e^{-\Phi_j p^{\theta_j}} \end{aligned}$$

where $\Phi_j = \sum_{ik} T_{ijk}\left(t_{kj}\tau_{ik}w_i\right)^{-\theta_j}$. If $p_{ij,k}(\nu) = p$ then the probability that ijk is the cost minimizing route-source choice is given by,

$$\begin{aligned} \prod_{i'\neq i,k'\neq k} \Pr\left[P_{ijrm} \geq p\right] &= \prod_{i'\neq i,k'\neq k}\left[1 - G_{i'jk'}\right] \\ &= \prod_{i'\neq i,k'\neq k} e^{-T_{ijk}\left(t_{kj}\tau_{ik}w_i\right)^{-\theta_j}p^{\theta_j}} \\ &= e^{-\left(\sum_{i,k} T_{ijr}\left(t_{kj}\tau_{ik}w_i\right)^{-\theta_j}\right)p^{\theta_j}} \end{aligned}$$

Integrating over all possible equilibrium prices p we can characterize the probability that node k and producer location i are the cost minimizing route-source choices:

$$\begin{aligned} \pi_{ij,k} &= \int_0^\infty \prod_{i'\neq i,k'\neq k}\left[1 - G_{i'jk'}\right] dG_{ijk}(p) \\ &= \int_0^\infty \prod_{i'\neq i,k'\neq k} e^{-T_{i'jk'}\left(t_{k'j'}\tau_{i'k'}w_{i'}\right)^{-\theta_j}p^{\theta_j}} dG_{ijk}(p) \end{aligned}$$

Replacing with $dG_{ijk}(p) = \left[T_{ijk}\left(t_{kj}\tau_{ij}w_i\right)^{-\theta_j}\theta_j p^{\theta_j - 1}\right] e^{-T_{ijk}\left(t_{kj}\tau_{ij}w_i\right)^{-\theta_j}p^{\theta_j}} dp$, we obtain,

$$\begin{aligned} \pi_{ij,k} &= \int_0^\infty \prod_{i'\neq i,k'\neq k} e^{-T_{i'jk'}\left(t_{k'j}\tau_{i'j}w_{i'}\right)^{-\theta_j}p^{\theta_j}} dG_{ijk}(p) \\ &= \int_0^\infty \prod_{i'\neq i,k'\neq k} e^{-T_{i'jk'}\left(t_{k'j}\tau_{i'j}w_{i'}\right)^{-\theta_j}p^{\theta_j}} \left[T_{ijk}\left(t_{kj}\tau_{ij}w_i\right)^{-\theta_j}\theta_j p^{\theta_j - 1}\right] e^{-T_{ijk}\left(t_{kj}\tau_{ij}w_i\right)^{-\theta_j}p^{\theta_j}} dp \\ &= T_{ijk}\left(t_{kj}\tau_{ij}w_i\right)^{-\theta_j} \int_0^\infty \prod_{i,k} e^{-T_{ijk}\left(t_{jk}\tau_{ji}w_i\right)^{-\theta_j}p^{\theta_j}} \left[\theta_j p^{\theta_j - 1}\right] dp \\ &= T_{ijk}\left(t_{kj}\tau_{ij}w_i\right)^{-\theta_j} \int_0^\infty e^{-\left(\sum_{i,k} T_{ijk}\left(t_{kj}\tau_{ij}w_i\right)^{-\theta_j}\right)p^{\theta_j}} \left[\theta_j p^{\theta_j - 1}\right] dp \\ &= T_{ijk}\left(t_{kj}\tau_{ij}w_i\right)^{-\theta_j} \int_0^\infty e^{-\Phi_j p^{\theta_j}} \left[\theta_j p^{\theta_j - 1}\right] dp \\ &= T_{ijk}\left(t_{kj}\tau_{ij}w_i\right)^{-\theta_j} \left[\frac{1}{\Phi_j} e^{-\Phi_j p^{\theta_j}}\right]_0^\infty \\ &= \frac{T_{ijk}\left(t_{kj}\tau_{ij}w_i\right)^{-\theta_j}}{\Phi_j} \end{aligned}$$

Replacing with $\Phi_j = \sum_{ik} T_{ijk} \left(t_{kj}\tau_{ij}w_i\right)^{-\theta_j}$, we obtain,

$$\pi_{ij,k} = \frac{T_{ijk}\left(t_{kj}\tau_{ij}w_i\right)^{-\theta_j}}{\sum_{ik} T_{ijk}\left(t_{kj}\tau_{ij}w_i\right)^{-\theta_j}}$$

replacing $T_{ijk} \equiv \left(\frac{1}{A_i}\right)^{\theta_j}$, we obtain,

$$\pi_{ij,k} = \frac{(w_i/A_i)^{-\theta_j}\left(t_{kj}\tau_{ik}\right)^{-\theta_j}}{\sum_{i\in\mathcal{N}}(w_i/A_i)^{-\theta_j}\sum_{k'\in\mathcal{F}(i)}\left(t_{k'j}\tau_{ik'}\right)^{-\theta_j}}$$

as stated above. Furthermore, the expected minimal price is given by,

$$\begin{aligned} p_{ij,k} &= \mathbb{E}\left[\min_{k\in\mathcal{F}(i)}\left\{\frac{t_{kj}\tau_{ik}w_i}{\varepsilon_{ij,k}(v)}\right\}\right] \\ &\propto \sum_{i\in\mathcal{N}}(w_i/A_i)^{-\theta_j}\sum_{k'\in\mathcal{B}(i)}\left(t_{k'j}\tau_{ik'}\right)^{-\theta_j} \\ &= \sum_{i\in\mathcal{N}}(w_i/A_i)^{-\theta_j}\,\tau_{ij}^{-\theta_j} \end{aligned}$$

where in the last line we have used the result that if we consider separately the routing problem this implies an expected transport cost that is in itself recursively defined (see Equation (3.1), Section 3.2),

$$\tau_{ij} = \mathbb{E}\left[\min_{k\in\mathcal{F}(i)}\left\{\frac{t_{ik}\tau_{kj}}{\varepsilon_{kj}}\right\}\right] \propto \left(\sum_{k\in\mathcal{B}(i)}\left(t_{ik}\tau_{kj}\right)^{-\theta_j}\right)^{-\frac{1}{\theta_j}}$$

which expressed the transport cost as an index of the continuation values along the different edges of the graph.

A.2 Derivations for Section 3.3: Combined Routing and Sourcing Choice, and Gravity

We follow the derivations for the general equilibrium gravity equation in Anderson and van Wincoop (2003) to derive a gravity formula, but allowing for generalized substitution patterns. Let us start by examining the expenditure shares, which are given by,

$$\pi_{ij} = \frac{\delta_i^{-\theta_j}\tau_{ij}^{-\theta_j}}{\sum_{i\in N}\delta_i^{-\theta_j}\tau_{ij}^{-\theta_j}} = \frac{\delta_i^{-\theta_j}\tau_{ij}^{-\theta_j}}{P_j^{\theta_j}}$$

where we have posited the existence of δ_i which is a generic origin-shifting gravity constant that is to be solved for. Combining this sourcing share expression to express bilateral trade flows, we obtain,

$$X_{ij} = \left(\frac{\delta_i\tau_{ji}}{P_j}\right)^{-\theta_j} Y_j$$

where we define the price index, allowing for node-specific substitution elasticities θ_j, i.e.

$$P_j = \left[\sum_{i\in N}\delta_i^{-\theta_j}\tau_{ij}^{-\theta_j}\right]^{-\frac{1}{\theta_j}}$$

We now impose market clearing to derive the expression for the gravity constant, i.e.

$$\begin{aligned}
Y_i &= \sum_j X_{ij} \\
&= \sum_j \left(\frac{\delta_i \tau_{ij}}{P_j} \right)^{-\theta_j} Y_j \\
&= (\delta_i)^{-\theta_i} \sum_j \delta_i^{\left(\theta_i - \theta_j\right)} P_j^{\theta_j} \tau_{ij}^{-\theta_j} Y_j \\
&= (\delta_i)^{-\theta_i} \sum_j \delta_i^{\Delta\theta_j} P_j^{\theta_j} \tau_{ij}^{-\theta_j} Y_j
\end{aligned}$$

where we defined the difference between the substitution parameters between nodes i and j as $\Delta\theta_j \equiv \theta_i - \theta_j$. Solving for δ_i,

$$\delta_i = \left(\frac{Y_i}{\sum_j \delta_i^{\Delta\theta_j} \left(\frac{\tau_{ij}}{P_j} \right)^{-\theta_j} Y_j} \right)^{-\frac{1}{\theta_i}}$$

Plugging δ_i back into the gravity equation, we obtain

$$\begin{aligned}
X_{ij} &= \left(\frac{\delta_i \tau_{ij}}{P_j} \right)^{-\theta_j} Y_j = (\delta_i)^{-\theta_j} \left(\frac{\tau_{ij}}{P_j} \right)^{-\theta_j} Y_j \\
&= \left(\frac{Y_i}{\sum_j \delta_i^{\Delta\theta_j} \left(\frac{\tau_{ij}}{P_j} \right)^{-\theta_j} Y_j} \right)^{\frac{\theta_j}{\theta_i}} \left(\frac{\tau_{ij}}{P_j} \right)^{-\theta_j} Y_j \\
&= \left(\frac{Y_i Y_j^{\frac{\theta_i}{\theta_j}}}{y^W} \right)^{\frac{\theta_j}{\theta_i}} \left(\frac{1}{\sum_j \delta_i^{\Delta\theta_j} \left(\frac{\tau_{ij}}{P_j} \right)^{-\theta_j} \frac{Y_j}{y^W}} \right)^{\frac{\theta_j}{\theta_i}} \left(\frac{\tau_{ij}}{P_j} \right)^{-\theta_j} \\
&= \left(\frac{Y_i Y_j^{\frac{\theta_i}{\theta_j}}}{y^W} \right)^{\frac{\theta_j}{\theta_i}} \left(\frac{1}{\left[\sum_j \delta_i^{\Delta\theta_j} \left(\frac{\tau_{ij}}{P_j} \right)^{-\theta_j} \frac{Y_j}{y^W} \right]^{-\frac{1}{\theta_j}}} \right)^{-\theta_j \frac{\theta_j}{\theta_i}} \left(\frac{\tau_{ij}}{P_j} \right)^{-\theta_j} \\
&= \left(\frac{Y_i Y_j^{\frac{\theta_i}{\theta_j}}}{y^W} \right)^{\frac{\theta_j}{\theta_i}} \left(\frac{1}{\left[\sum_j \delta_i^{\Delta\theta_j} \left(\frac{\tau_{ij}}{P_j} \right)^{-\theta_j} \frac{Y_j}{y^W} \right]^{-\frac{1}{\theta_j}}} \right)^{-\theta_j \frac{\theta_j}{\theta_i}} \left(\frac{\tau_{ij}}{P_j} \right)^{-\theta_j} \\
&= \left(\frac{Y_i Y_j^{\frac{\theta_i}{\theta_j}}}{y^W} \right)^{\frac{\theta_j}{\theta_i}} \left(\frac{\left(\tau_{ij}\right)^{\frac{\theta_i}{\theta_j}}}{\left[\sum_j \delta_i^{\Delta\theta_j} \left(\frac{\tau_{ij}}{P_j} \right)^{-\theta_j} \frac{Y_j}{y^W} \right]^{-\frac{1}{\theta_j}} P_j^{\frac{\theta_i}{\theta_j}}} \right)^{-\theta_j \frac{\theta_j}{\theta_i}} \\
&= \left(\frac{Y_i Y_j^{\frac{\theta_i}{\theta_j}}}{y^W} \right)^{\frac{\theta_j}{\theta_i}} \left(\frac{\left(\tau_{ij}\right)^{\frac{\theta_i}{\theta_j}}}{\Pi_i P_j^{\frac{\theta_i}{\theta_j}}} \right)^{-\theta_j \frac{\theta_j}{\theta_i}}
\end{aligned}$$

where the producer index is given by,

$$\Pi_i = \left(\sum_j \delta_i^{\Delta\theta_j} \left(\frac{\tau_{ij}}{P_j} \right)^{-\theta_j} \frac{Y_j}{y^W} \right)^{-\frac{1}{\theta_j}}$$

We can substitute in the equilibrium scaled prices to get,

$$\begin{aligned}
P_j &= \left[\sum_i (\delta_i)^{-\theta_j} \left(\tau_{ij}\right)^{-\theta_j} \right]^{-\frac{1}{\theta_j}} \\
&= \left[\sum_i \left(\frac{Y_i}{\sum_j \delta_i^{\Delta\theta_j} \left(\frac{\tau_{ij}}{P_j} \right)^{-\theta_j} y_j} \right)^{\frac{\theta_j}{\theta_i}} \left(\tau_{ij}\right)^{-\theta_j} \right]^{-\frac{1}{\theta_j}} \\
&= \left[\sum_i \left(\frac{\frac{Y_i}{y^W}}{\sum_j \delta_i^{\Delta\theta_j} \left(\frac{\tau_{ij}}{P_j} \right)^{-\theta_j} \frac{Y_j}{y^W}} \right)^{\frac{\theta_j}{\theta_i}} \left(\tau_{ij}\right)^{-\theta_j} \right]^{-\frac{1}{\theta_j}} \\
&= \left[\sum_i \left(\frac{Y_i}{y^W} \right)^{\frac{\theta_j}{\theta_i}} \left(\frac{\left(\tau_{ij}\right)^{\frac{\theta_i}{\theta_j}}}{\left(\sum_j \delta_i^{\Delta\theta_j} \left(\frac{\tau_{ij}}{P_j} \right)^{-\theta_j} \frac{Y_j}{y^W} \right)^{-\frac{1}{\theta_j}}} \right)^{-\theta_j \frac{\theta_j}{\theta_i}} \right]^{-\frac{1}{\theta_j}} \\
&= \left[\sum_i \tau_{ij}^{-\theta_j} \left(\frac{\frac{Y_i}{y^W}}{\Pi_i} \right)^{\frac{\theta_j}{\theta_i}} \right]^{-\frac{1}{\theta_j}}
\end{aligned}$$

In conclusion we have the following producer and consumer price indices respectively,

$$\Pi_i \equiv \left(\sum_j \delta_i^{\Delta\theta_j} \left(\frac{\tau_{ij}}{P_j} \right)^{-\theta_j} \gamma_j \right)^{-\frac{1}{\theta_j}}$$

$$P_j = \left[\sum_i \tau_{ij}^{-\theta_j} \left(\frac{\gamma_i}{\Pi_i} \right)^{\frac{\theta_j}{\theta_i}} \right]^{-\frac{1}{\theta_j}}$$

and we obtain the following gravity equation,

$$X_{ij} = \tau_{ij}^{-\theta_j} \times \left(\frac{\gamma_i}{\Pi_i^{-\theta_j}} \right)^{\frac{\theta_j}{\theta_i}} \times \frac{\delta_j}{P_j^{-\theta_j}} \tag{A.1}$$

where $\delta_i = Y_i$ and $\gamma_i = \frac{Y_j}{y^w}$ represent the gravity constants, P_j is the consumer price index or inward market access term, and Π_i is the producer price index or outward market access term.[1] For symmetric substitution parameters (if $\theta_j = \theta_i = \theta$), both the gravity formula and price indices collapse to their traditional form.

[1] Producer price index is $\Pi_i \equiv \left(\sum_j \delta_i^{(\theta_i - \theta_j)} \left(\frac{\tau_{ij}}{P_j} \right)^{-\theta_j} \gamma_j \right)^{-\frac{1}{\theta_j}}$ while consumer price index is $P_j = \left[\sum_i \tau_{ij}^{-\theta_j} \left(\frac{\gamma_i}{\Pi_i} \right)^{\frac{\theta_j}{\theta_i}} \right]^{-\frac{1}{\theta_j}}$.

A.3 Derivations for Section 3.3: Combined Routing and Sourcing Choice, and Recursive Equilibrium

This section provides additional derivations for the equilibrium equations.

In terms of market clearing and trade balance we have the following two equilibrium conditions,

$$Y_i = \sum_j X_{ij}$$
$$E_i = \sum_j X_{ji}$$

Starting with the first equilibrium condition we can derive the first part of Equation (3.4):

$$\gamma_i = \sum_j X_{ij}$$
$$\gamma_i = \sum_j \tau_{ij}^{-\theta_j} \times \left(\frac{\gamma_i}{\Pi_i^{-\theta_j}} \right)^{\frac{\theta_j}{\theta_i}} \times \frac{\delta_j}{P_j^{-\theta_j}}$$
$$\Pi_i^{-\theta_i} = \sum_j \tau_{ij}^{-\theta_j} \times \Pi_i^{\theta_j \frac{\theta_j}{\theta_i} - \theta_i} (\gamma_i)^{\frac{\theta_j}{\theta_i} - 1} \times \frac{\delta_j}{P_j^{-\theta_j}}$$
$$\Pi_i^{-\theta_i} = \sum_j \tau_{ij}^{-\theta_j} \times \Pi_i^{\Delta_{i,j}} \times (\gamma_i)^{\frac{\theta_j}{\theta_i} - 1} \times \frac{\delta_j}{P_j^{-\theta_j}}$$
$$\Pi_i^{-\theta_i} = \tau_{ii}^{-\theta_i} \frac{\delta_i}{P_i^{-\theta_i}} + \sum_{j \neq i} \tau_{ij}^{-\theta_j} \Pi_i^{\Delta_{i,j}} (\gamma_i)^{\frac{\theta_j}{\theta_i} - 1} \frac{\delta_j}{P_j^{-\theta_j}}$$
$$\Pi_i^{-\theta_i} = \tau_{ii}^{-\theta_i} \frac{\delta_i}{P_i^{-\theta_i}} + \sum_{j \neq i} \left(\sum_{k \in \mathscr{B}(k)} t_{ik}^{-\theta_j} \tau_{kj}^{-\theta_j} \right) \Pi_i^{\Delta_{i,j}} (\gamma_i)^{\frac{\theta_j}{\theta_i} - 1} \frac{\delta_j}{P_j^{-\theta_j}}$$
$$\Pi_i^{-\theta_i} = \tau_{ii}^{-\theta_i} \frac{\delta_i}{P_i^{-\theta_i}} + \sum_{j \neq i} \left(\sum_{k \in \mathscr{F}(i)} t_{ik}^{-\theta_j} \tau_{kj}^{-\theta_j} \right) \Pi_i^{\Delta_{i,j}} (\gamma_i)^{\frac{\theta_j}{\theta_i} - 1} \frac{\delta_j}{P_j^{-\theta_j}}$$
$$\Pi_i^{-\theta_i} = \tau_{ii}^{-\theta_i} \frac{\delta_i}{P_i^{-\theta_i}} + \sum_{k \in \mathscr{F}(i)} t_{ik}^{-\theta_i} \left(-\tau_{ki}^{-\theta_i} \frac{\delta_i}{p_i^{-\theta_i}} + \sum_j t_{ik}^{\Delta\theta_j} \tau_{kj}^{-\theta_j} \Pi_i^{\Delta_{i,j}} (\gamma_i)^{\frac{\theta_j}{\theta_i} - 1} \frac{\delta_j}{P_j^{-\theta_j}} \right)$$
$$\Pi_i^{-\theta_i} = \left(t_{ii}^{-\theta_i} + \sum_{k \in \mathscr{F}(i)} (t_{ik} \tau_{ki})^{-\theta_i} \right) \frac{\delta_i}{P_i^{-\theta_i}} + \sum_{k \in \mathscr{F}(i)} t_{ik}^{-\theta_i} \left(-\tau_{ki}^{-\theta_i} \frac{\delta_i}{p_i^{-\theta_i}} + \sum_j t_{ik}^{\Delta\theta_j} \tau_{kj}^{-\theta_j} \Pi_i^{\Delta_{i,j}} (\gamma_i)^{\frac{\theta_j}{\theta_i} - 1} \frac{\delta_j}{P_j^{-\theta_j}} \right)$$
$$\Pi_i^{-\theta_i} = \left(t_{ii}^{-\theta_i} \right) \frac{\delta_i}{P_i^{-\theta_i}} + \sum_{k \in \mathscr{F}(i)} t_{ik}^{-\theta_i} \left(\sum_j t_{ik}^{\Delta\theta_j} \Pi_i^{\Delta_{i,j}} (\gamma_i)^{\frac{\theta_j}{\theta_i} - 1} \tau_{kj}^{-\theta_j} \frac{\delta_j}{P_j^{-\theta_j}} \right)$$
$$\Pi_i^{-\theta_i} = t_{ii}^{-\theta_i} \frac{\delta_i}{P_i^{-\theta_i}} + \sum_{k \in \mathscr{F}(i)} t_{ik}^{-\theta_i} \tilde{\Pi}_k^{-\theta_i}$$

where in the second line we defined, $\Pi_i^{\Delta_{i,j}} \equiv \Pi_i^{\theta_j \frac{\theta_j}{\theta_i} - \theta_i}$ and where $\tilde{\Pi}_k^{-\theta_i} = \left(\sum_j t_{ik}^{\Delta\theta_j} \Pi_i^{\Delta_{i,j}} (\gamma_i)^{\frac{\theta_j}{\theta_i} - 1} \tau_{kj}^{-\theta_j} \frac{\delta_j}{P_j^{-\theta_j}} \right)^{\frac{\theta_i}{\theta_j}}$ is the approximate price index that controls for differences in elasticities of substitution between neighboring nodes. And where in the fourth line we have used the fact that at the penultimate node the set of predecessor nodes of k ($\mathscr{B}(k)$) coincides with set of successor nodes of i ($\mathscr{F}(i)$). And finally, where in the last line we used the definition of the recursive transport cost (Equation (3.1)), $\tau_{ii}^{-\theta} = \left(t_{ii}^{-\theta} + \sum_{k \in \mathscr{F}(i)} \left(t_{ik} \tau_{kj} \right)^{-\theta} \right)$.

Continuing with the second equilibrium condition we can derive the second part of Equation (3.4),

$$\delta_j = \sum_i X_{ij}$$

$$\delta_j = \sum_i \tau_{ij}^{-\theta_j} \times \left(\frac{\gamma_i}{\Pi_i^{-\theta_j}} \right)^{\frac{\theta_j}{\theta_i}} \times \frac{\delta_j}{P_j^{-\theta_j}}$$

$$P_j^{-\theta_j} = \sum_i \tau_{ij}^{-\theta_j} \times \left(\frac{\gamma_i}{\Pi_i^{-\theta_j}} \right)^{\frac{\theta_j}{\theta_i}}$$

$$P_j^{-\theta_j} = \tau_{jj}^{-\theta_j} \frac{\gamma_j}{\Pi_j^{-\theta_j}} + \sum_{i \neq j} \tau_{ij}^{-\theta_j} \left(\frac{\gamma_i}{\Pi_i^{-\theta_j}} \right)^{\frac{\theta_j}{\theta_i}}$$

$$P_j^{-\theta_j} = \tau_{jj}^{-\theta_j} \frac{\gamma_j}{\Pi_j^{-\theta_j}} + \sum_{i \neq j} \left(\sum_{k \in \mathscr{B}(j)} \tau_{ik}^{-\theta_j} t_{kj}^{-\theta_j} \right) \left(\frac{\gamma_i}{\Pi_i^{-\theta_j}} \right)^{\frac{\theta_j}{\theta_i}}$$

$$P_j^{-\theta_j} = \tau_{jj}^{-\theta_j} \frac{\gamma_j}{\Pi_j^{-\theta}} + \sum_{k \in \mathscr{B}(j)} t_{kj}^{-\theta_j} \left(-\tau_{jk}^{-\theta_j} \frac{\gamma_j}{\Pi_j^{-\theta_j}} + \sum_i \tau_{ik}^{-\theta_j} \left(\frac{\gamma_i}{\Pi_i^{-\theta_j}} \right)^{\frac{\theta_j}{\theta_i}} \right)$$

$$P_j^{-\theta_j} = \left(t_{jj}^{-\theta_j} + \sum_{k \in \mathscr{B}(j)} \left(t_{jk} \tau_{kj} \right)^{-\theta_j} \right) \frac{\gamma_j}{\Pi_j^{-\theta_j}} + \sum_{k \in \mathscr{B}(j)} t_{kj}^{-\theta_j} \left(-\tau_{jk}^{-\theta_j} \frac{\gamma_j}{\Pi_j^{-\theta_j}} + \sum_i \tau_{ik}^{-\theta_j} \left(\frac{\gamma_i}{\Pi_i^{-\theta_j}} \right)^{\frac{\theta_j}{\theta_i}} \right)$$

$$P_j^{-\theta_j} = \left(t_{jj}^{-\theta_j} \right) \frac{\gamma_j}{\Pi_j^{-\theta_j}} + \sum_{k \in \mathscr{B}(j)} t_{kj}^{-\theta_j} \left(\sum_i \tau_{ik}^{-\theta_j} \left(\frac{\gamma_i}{\Pi_i^{-\theta_j}} \right)^{\frac{\theta_j}{\theta_i}} \right)$$

$$P_j^{-\theta_j} = t_{jj}^{-\theta_j} \frac{\gamma_j}{\Pi_j^{-\theta_j}} + \sum_{k \in \mathscr{B}(j)} t_{kj}^{-\theta_j} P_k^{-\theta_j}$$

where in the last line we used the definition of the recursive transport cost (Equation (3.1)), $\tau_{ii}^{-\theta} = \left(t_{ii}^{-\theta} + \sum_{k \in \mathscr{F}(i)} \left(t_{ik} \tau_{kj} \right)^{-\theta} \right)$. Therefore we have,

$$P_j^{-\theta_j} = t_{jj}^{-\theta_j} \frac{\gamma_j}{\Pi_j^{-\theta_j}} + \sum_{k \in \mathscr{B}(j)} t_{kj}^{-\theta_j} P_k^{-\theta_j}$$

$$\Pi_i^{-\theta_i} = \left(t_{ii}^{-\theta_i} \right) \frac{\delta_i}{P_i^{-\theta_i}} + \sum_{k \in \mathscr{F}(i)} t_{ik}^{-\theta_i} \left(\sum_j t_{ik}^{\Delta \theta_j} \Pi_i^{\Delta_{i,j}} (\gamma_i)^{\frac{\theta_j}{\theta_i} - 1} \tau_{kj}^{-\theta_j} \frac{\delta_j}{P_j^{-\theta_j}} \right)$$

Notice that for the case where we have $\theta_i = \theta_j$, this reduces to,

$$\Pi_i^{-\theta} = \left(t_{ii}^{-\theta} \right) \frac{\delta_i}{P_i^{-\theta}} + \sum_{k \in \mathscr{F}(i)} t_{ik}^{-\theta} \left(\sum_j \tau_{kj}^{-\theta} \frac{\delta_j}{P_j^{-\theta}} \right) = \left(t_{ii}^{-\theta} \right) \frac{\delta_i}{P_i^{-\theta}} + \sum_{k \in \mathscr{F}(i)} t_{ik}^{-\theta} \Pi_k^{-\theta}$$

$$P_j^{-\theta} = t_{jj}^{-\theta} \frac{\gamma_j}{\Pi_j^{-\theta}} + \sum_{k \in \mathscr{B}(j)} t_{kj}^{-\theta} P_k^{-\theta}$$

We can further express the price indices in terms of equilibrium labor and income allocations across space. We assume welfare equalization, $W_j = \frac{w_j}{P_j} u_j$, and assume localized productivity (A_i) and amenity spillovers (u_i) that depend on the density of workers in a locality, $A_i = \bar{A}_i L_i^{\alpha}$ and $u_i = \bar{u}_i L_i^{\beta}$, where $\bar{A}_i$ is the exogenous productivity component at location i and α determines the extent to which productivity is affected by the local population L_i (productivity spillovers), and where $\bar{u}_i$ is the exogenous utility derived from living in location i and β governs the extent to which amenities are affected by the location population (amenity spillovers). We rewrite the producer and consumer price index respectively,

$$P_i = \frac{w_i u_i}{W} \iff$$
$$P_i = Y_i \bar{u}_i L_i^{\beta-1} W^{-1}$$

$$\Pi_i = A_i L_i Y_i^{-\frac{\theta+1}{\theta}} \iff$$
$$\Pi_i = A_i L_i^{\alpha+1} Y_i^{-\frac{\theta+1}{\theta}}$$

As a result, we have $P_j = \frac{1}{W} \bar{u}_j L_j^{\beta-1} Y_j$ and $\Pi_i = \bar{A}_i L_i^{1+\alpha} Y_i^{-\frac{\theta_i+1}{\theta_i}}$.[2] Specifically, given the productivity and amenity fundamentals $\left\{\bar{A}_i, \bar{u}_i\right\}$ as well as the aggregate transport costs, the system of $2N$ equations can be solved for the $2N$ endogenous equilibrium values. The equilibrium system determines the endogenous variables via the interaction of price indices along the network topology, where transport cost is endogenously determined as part of the equilibrium system.

A.4 Derivations for Section 3.4: Nested Mode Choice and Aggregate Transport Cost

As presented in the previous section, we consider a consumer that resides in location j and makes a route-sourcing choice by choosing sequentially edges along the graph. This provides a convenient characterization of routing that avoids the curse of dimensionality by expressing the problem as a recursive problem instead of considering the universe of possible routes along a possibly high-dimensional graph. To furthermore accommodate multimodal routing choices, we incorporate a nested modal choice. In that setting, conditional on the neighboring node chosen, the consumer makes a modal choice by choosing the cost minimizing mode out of all modes along the edge, i. to traverse the edge (subject to an extreme value distributed cost shock). To fix ideas, consider a consumer in location j having chosen to route towards the neighboring node k. The consumer then compares all the different modes that are available along this edge, i.e. $m \in \mathscr{M}_{jk}$, where $\mathscr{M}_{jk}$ is the set of modes available between nodes $j$and k, and where the edge mode specific transport costs given by,

$$\tilde{t}_{jk,m} = \begin{cases} t_{jk,m} & \text{if} \quad m = 1 \\ s_{jj,m} \tau_{jk,m} s_{kk,m} & \text{if} \quad m \neq 1 \end{cases}$$

where for the primary mode no switching cost is required, but for any non-primary mode ($m \neq 1$) a switching cost ist imposed, while $\tau_{jk,m}$ refers to the iceberg transport cost of traversing the edge between node j and k along mode m. Notice that this specification is extremeley general and allows for geographies where non-primary modes might connect an entirely different set of nodes than primary nodes. The consumer then faces a cost minimizing choice subject to an extreme-value distributed cost shock, i.e.

$$\min_{m \in \mathscr{M}(j,k)} \left\{ \frac{t_{jk}}{\varepsilon_m} \right\}$$

where the properties of the Frechet distribution implies that the expected cost minimizing transport along any mode between nodes j and k is given by,

$$t_{ik}^{-\theta} \propto \left(\sum_{m \in \mathscr{M}(i,k)} \tilde{t}_{ik,m}^{-\eta} \right)^{\frac{\theta}{\eta}|}$$

Overall, the consumer's route-sourcing choice can be written as a nested minimization problem where we can characterize

[2] We can furthermore rewrite the equations in terms of rescaled variables, $\{y_i, l_i\}$, where we define shares of world income in location i, $y_i \equiv \frac{Y_i}{Y^W}$, and shares of total labor in location i, $l_i \equiv \frac{L_i}{L^W}$. To illustrate this, we did this explicitly for the counterfactual equilibrium system in the Online Appendix A.10.

the overall expected minimal cost, i.e.

$$\begin{aligned}
p_{ij} &= \mathbb{E}\left[\min_{j\in\mathscr{F}(i)}\left\{\mathbb{E}\left[\min_{m\in\mathscr{M}(i,k)}\left\{\frac{t_{ik,m}}{\varepsilon_m}\right\}\right]\tau_{jd}w_i\frac{1}{\varepsilon_{ij}}\right\}\right] \\
&\propto \sum_{i\in\mathscr{N}}(w_i/A_i)^{-\theta}\sum_{k'\in\mathscr{F}(i)}\left(\left(\sum_{m\in\mathscr{M}(i,k)}\tilde{t}_{ik,m}^{-\eta}\right)^{-\frac{1}{\eta}}\tau_{k'i}\right)^{-\theta} \\
&= \sum_{i\in\mathscr{N}}(w_i/A_i)^{-\theta}\,\tau_{ij}^{-\theta}
\end{aligned}$$

where in the last line we again have used the definition of the transportation cost in terms of the edge-specific costs and associated continuation values along neighboring nodes.

$$\tau_{id} = \mathbb{E}\left[\min_{j\in\mathscr{F}(i)}\left\{\frac{t_{ij}\tau_{jd}}{\varepsilon_{ij}}\right\}\right] \propto \left(\sum_{j\in\mathscr{F}(i)}\left(\left(\sum_{m\in\mathscr{M}(i,k)}t_{ik,m}^{-\eta}\right)^{-\frac{1}{\eta}|}\tau_{jd}\right)^{-\theta}\right)^{-\frac{1}{\theta}}$$

which expressed the transport cost as an index of the continuation values along the different edges of the graph.

The mode-route-sourcing choice probability is therefore given by,

$$\begin{aligned}
\pi_{ij,k,m} &= \frac{(w_i/A_i)^{-\theta}\,(t_{ik}\tau_{ki})^{-\theta}}{\sum_{i\in\mathscr{N}}(w_i/A_i)^{-\theta}\sum_{k'\in\mathscr{F}(i)}(t_{ik'}\tau_{k'i})^{-\theta}}\frac{\tilde{t}_{ik,m}^{-\eta}}{t_{ik}^{-\eta}} \\
&\equiv \pi_{ij,k}\times\pi_{ik}^{m}
\end{aligned}$$

where in the last line, we can decompose this probability into the link choice probability ($\pi_{ij,k}$) and the mode choice probability along the link (π_{ik}^{m})

A.5 Isomorphism with Allen and Arkolakis (2022)

It might be insightful to link the recursive expression in our paper to the approach in AA2022 that instead relies on explicit enumeration of the universe of paths and utilizes matrix algebra to express the expected minimum transport cost in terms of the leontief inverse of the adjacency matrix that captures the underlying infrastructure network. As we will show, the model is isomorphic in the case where there is a simple unimodal transport network. Furthermore, we will also show that for the simplified case where the mode choice elasticity is equal to the route choice elasticity, we can show that their approach can be extended to provide a clean expression of the multimodal transport cost in terms of (leontief inverses) of the infrastructure matrices. First, let us demonstrate the isomorphism for unimodal transport network. Consider the recursive transport cost stated above,

$$\tau_{id} = \mathbb{E}\left[\min_{j\in\mathscr{F}(i)}\left\{\frac{t_{ij}\tau_{jd}}{\varepsilon_{ij}}\right\}\right] = \left(\sum_{j\in\mathscr{F}(i)}\left(t_{ij}\tau_{jd}\right)^{-\theta}\right)^{-\frac{1}{\theta}}$$

If we assume a finite graph, then we can iteratively substitute and obtain a closed-form expression for the endogenous transport cost, i.e.

$$\begin{aligned}
(\tau_{id})^{-\theta} &= t_{d-1d}^{-\theta}\sum_{j\in F(i)}t_{ij}^{-\theta}\sum_{j'\in F(j)}\cdots\sum_{k'\in F(d-1)}t_{k'd-1}^{-\theta} \\
&= \sum_{r\in\mathfrak{R}_{ij}}\left(\prod_{l=1}^{K}t_{r_{l-1},r_l}^{-\theta}\right) \\
&= \sum_{K=0}^{\infty}A_{ij}^{K} \\
&= (\mathbf{I}-\mathbf{A})^{-1}\equiv\mathbf{B}
\end{aligned}$$

where in the second line we recognize that the recursive substitution on a finite graph results in a characterization of all possible routes of any length along the network. The resulting expression is identical to expression for endogenous transport cost in AA2022. In the third and fourth line we then employ the same argument as in their paper to show that the implied transport cost can be expressed as the leontief inverse of the underlying infrastructure matrix, where $\mathbf{A} \equiv \left[t_{ij}^{-\theta} \right]$ is an $N \times N$ matrix with (i, j) element $t_{ij}^{-\theta}$ and A_{ij}^{K} is the (ij) element of the matrix A to the matrix power K. This shows that in essence, the two different approaches are isomorphic and capture the same underlying endogenous transport cost, albeit in different ways.

We now turn towards showing that for the special case where the mode choice elasticity is equal to the route choice elasticity, we can furthermore extend the approach in AA2022 to derive a clean decomposition of the multimodal transport cost in terms of a set of underlying mode-specific infrastructure matrices. Let us therefore consider the case where $\eta = \theta$. Notice that in this special case, the edge-specific transport cost is given by,

$$t_{ik}^{-\theta} \propto \sum_{m \in \mathscr{M}(i,k)} \tilde{t}_{ik,m}^{-\theta}$$

where $\tilde{t}_{ik,m}^{-\theta}$ are the mode specific traversal costs between node i and k. As above, we can express recursively substitute the expected transport across modes between nodes, and obtain,

$$(\tau_{id})^{-\theta} = t_{d-1d}^{-\theta} \sum_{j \in F(i)} t_{ij}^{-\theta} \sum_{j' \in F(j)} \dots \sum_{k' \in F(d-1)} t_{k'd-1}^{-\theta}$$

By substituting the edge-specific transport cost in terms of the mode-specific cost, we obtain,

$$\begin{aligned}(\tau_{id})^{-\theta} &= \left(\sum_{m \in \mathscr{M}(d-1,d)} \tilde{t}_{d-1d,m}^{-\theta} \right) \sum_{j \in F(i)} \left(\sum_{m \in \mathscr{M}(i,j)} \tilde{t}_{ij,m}^{-\theta} \right) \sum_{j' \in F(j)} \dots \sum_{k' \in F(d-1)} \left(\sum_{m \in \mathscr{M}(k',d-1)} \tilde{t}_{k'd-1,m}^{-\theta} \right) \\ &= \sum_{r \in \Re_{ij}} \left(\prod_{l=1}^{K} \sum_{m \in \mathscr{M}(r_{l-1}, r_l)} \tilde{t}_{r_{l-1}, r_l, m}^{-\theta} \right)\end{aligned}$$

where the second line is a concise way of summarizing all possible uni and multi-modal paths along the multi-layered network.Without loss of generality, let us consider the case of two modes - one primary mode where flows originate and terminate, and a secondary mode that is accessible subject to some switching cost. Define the $(N_1 + N_2) \times (N_1 + N_2)$ matrix $\mathbf{A} = \left[a_{ij} \equiv t_{ij}^{-\theta} \right]$. Notice that this adjacency matrix forms a block partitioned matrix, i.e.

$$\mathbf{A} = \begin{bmatrix} \mathbf{A_1} & \mathbf{S} \\ \mathbf{S}' & \mathbf{A_2} \end{bmatrix}$$

where $\mathbf{A_1} = \left[a_{ij} \right] = \left[t_{ij}^{-\theta} \right]$ is the adjacency matrix for the primary transportation network, $\mathbf{A_2} = \left[a_{i'j'} \right] = \left[t_{i'j'}^{-\theta} \right]$ is the adjacency matrix for the secondary transportation network, and $\mathbf{S} = \left[s_{ii'}^{-\theta} \right]$ is the diagonal matrix that represents linkages between the primary and secondary transportation network. We can write τ_{ij} from by explicitly summing across all possible routes of all possible lengths. To do so, we sum across all locations that are traveled through all the possible paths as follows:

$$\begin{aligned}\tau_{ij}^{-\theta} &= \sum_{r \in \Re_{ij}} \left(\prod_{l=1}^{K} \sum_{m \in \mathscr{M}(r_{l-1}, r_l)} \tilde{t}_{r_{l-1}, r_l, m}^{-\theta} \right) \\ &= \sum_{K=0}^{\infty} \left(\sum_{k_1=1}^{(N_1+N_2)} \sum_{k_2=1}^{(N_1+N_2)} \dots \sum_{k_{K-1}=1}^{(N_1+N_2)} a_{i,k_1} \times a_{k_1,k_2} \times \dots \times a_{k_{K-2},k_{K-1}} \times a_{k_{K-1}j} \right)\end{aligned}$$

By explicitly recognizing that this sum across all locations through all possible paths can be partitioned into unimodal paths on each transportation network and an arbitrary number of switches between transportation modes, we have,

$$\tau_{ij}^{-\theta} = \sum_{t_1=1}^{N} \sum_{t_2=1}^{N} \dots \sum_{t_S=1}^{N} \left(\left(\sum_{K=0}^{\infty} \mathbf{A}_{1,it_1}^{K} \right) \times s_{t_1 t_1'}^{-\theta} \times \dots \times s_{t_s' t_s}^{-\theta} \left(\sum_{K=0}^{\infty} \mathbf{A}_{1,T_S j}^{K} \right) \right)$$

which in matrix notation can be written as,

$$\tau_{ij}^{-\theta} = \sum_{K=0}^{\infty}\left(\left(\sum_{K=0}^{\infty}\mathbf{A}_1^K\right)\left(\mathbf{S}\left(\sum_{K=0}^{\infty}\mathbf{A}_2^K\right)\mathbf{S}'\right)\right)^K\left(\sum_{K=0}^{\infty}\mathbf{A}_1^K\right)$$

To simplify this expression let us first define the Leontief inverse for each infrastructure matrix separately, i.e.

$$\sum_{K=0}^{\infty}\mathbf{A}_1^K = (\mathbf{I}-\mathbf{A}_1)^{-1} \equiv \mathbf{B}$$

$$\sum_{K=0}^{\infty}\mathbf{A}_2^K = (\mathbf{I}-\mathbf{A}_2)^{-1} \equiv \mathbf{C}$$

We also define - for convenience - the sandwich matrix that adjusts the transport cost along the secondary transportation network for switching costs and therefore traces out the option value of having access to the secondary transportation network,

$$\mathbf{S}\left(\sum_{K=0}^{\infty}\mathbf{A}_2^K\right)\mathbf{S}' \equiv \mathbf{D}$$

From matrix calculus we can restate the following result that relates the inverse of the Schur complement of the partitioned infrastructure matrix to the geometric sum of matrix operations, specifically,

$$\sum_{K=0}^{\infty}\left(\mathbf{B}^{-1}\mathbf{D}\right)^K\mathbf{B}^{-1} = (\mathbf{B}-\mathbf{D})^{-1} \equiv \mathbf{E}$$

Applying this result, we can write,

$$\begin{aligned}\tau_{ij}^{-\theta} &= \sum_{K=0}^{\infty}\left((\mathbf{I}-\mathbf{A}_1)^{-1}\left(\mathbf{S}(\mathbf{I}-\mathbf{A}_2)^{-1}\mathbf{S}'\right)\right)^K(\mathbf{I}-\mathbf{A}_1)^{-1}\\ &= \left[(\mathbf{I}-\mathbf{A}_1)-\mathbf{S}(\mathbf{I}-\mathbf{A}_2)^{-1}\mathbf{S}'\right]^{-1}_{ij}\end{aligned}$$

therefore we can write,

$$\tau_{ij} = e_{ij}^{-\frac{1}{\theta}}$$

Furthermore, the Woodbury matrix identity (see e.g. Horn and Johnson (2012)) states,

$$(\mathbf{A}+\mathbf{UCV})^{-1} = \mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{U}\left(\mathbf{C}^{-1}+\mathbf{VA}^{-1}\mathbf{U}\right)^{-1}\mathbf{VA}^{-1}$$

which implies

$$\begin{aligned}\tau_{ij}^{-\theta} &= \left[(\mathbf{I}-\mathbf{A}_1)-\mathbf{S}(\mathbf{I}-\mathbf{A}_2)^{-1}\mathbf{S}'\right]^{-1}_{ij}\\ &= \left[\mathbf{B}+\mathbf{BS}(\mathbf{A}/\mathbf{A}_1)^{-1}\mathbf{S}'\mathbf{B}\right]_{ij}\\ &= \left[(\mathbf{I}-\mathbf{A}_1)^{-1}+(\mathbf{I}-\mathbf{A}_1)^{-1}\mathbf{S}(\mathbf{A}/\mathbf{A}_1)^{-1}\mathbf{S}'(\mathbf{I}-\mathbf{A}_1)^{-1}\right]_{ij}\end{aligned}$$

where $\mathbf{A}/\mathbf{A}_1 := (\mathbf{I}-\mathbf{A}_1)^{-1}-\mathbf{S}(\mathbf{A}/\mathbf{A}_1)^{-1}\mathbf{S}'$ defines the Schur complement of the adjacency matrix A. The expressions corresponds to the expression given in the main text and intuitively decomposes the transport cost into a component that originates from the unimodal paths and another component that originates from the multimodal paths. This result can also directly be obtained by applying to the partitioned matrix A the formula for the inverse of block-partitioned matrices (see e.g. Horn and Johnson (2012)).

A.6 Derivations for Section 3.5: Congestion and Traffic Flows

We characterize equilibrium traffic at different nodes of the transportation network. First, we utilize the recursive routing choice to characterize aggregate traffic between any two nodes across any mode. Second, we characterize mode-specific traffic between nodes. Finally, we characterize traffic at terminal stations.

A.6.1 Traffic Flows on the Aggregate Network

We start by characterizing the aggregate traffic between any two nodes across any mode. To do so we start by restating the (recursively defined) sourcing and link choice probability which is given by:

$$\pi_{ij,k} = \frac{(w_i/A_i)^{-\theta}\left(t_{jk}\tau_{ki}\right)^{-\theta}}{\sum_{i\in\mathcal{N}}(w_i/A_i)^{-\theta}\sum_{k'\in\mathcal{F}(i)}\left(t_{jk'}\tau_{k'i}\right)^{-\theta}}$$

which can be decomposed in the sourcing share and the link choice probability conditional on the sourcing choice, i.e.

$$\begin{aligned}
\pi_{ij,kl} &= \frac{(t_{ik}\tau_{ki})^{-\theta}}{\sum_{k'\in\mathcal{F}(i)}(t_{ik'}\tau_{k'i})^{-\theta}}\frac{\tau_{ij}^{-\theta}p_i^{-\theta}}{\sum_{i\in\mathcal{N}}\tau_{ij}^{-\theta}p_i^{-\theta}} \\
&= \frac{(t_{ik}\tau_{ki})^{-\theta}}{\tau_{ij}^{-\theta}}\frac{\tau_{ij}^{-\theta}p_i^{-\theta}}{\sum_{i\in\mathcal{N}}\tau_{ij}^{-\theta}p_i^{-\theta}} \\
&= \pi_{ij}^{k}\times\pi_{ij}
\end{aligned}$$

The previous derivations only characterize the probability of choosing neighboring link k when routing between i and j. Notice that we can construct the probability of using any edge kl when transporting goods from i to j in straightforward way,

$$\begin{aligned}
\pi_{kj}^{kl} &= \frac{t_{kl}^{-\theta}\tau_{lj}^{-\theta}}{\tau_{kj}^{-\theta}} \\
&= \frac{\tau_{ik}^{-\theta}t_{kl}^{-\theta}\tau_{lj}^{-\theta}}{\tau_{ik}^{-\theta}\tau_{kj}^{-\theta}} \\
&= \frac{\tau_{ik}^{-\theta}t_{kl}^{-\theta}\tau_{lj}^{-\theta}}{\tau_{ij}^{-\theta}} \\
&= \left(\frac{\tau_{ij}}{\tau_{ik}t_{kl}\tau_{ij}}\right)^{\theta}
\end{aligned}$$

which gives us the traffic equation in Section 3. Notice that this arises naturally due to the markovian property of the recursive routing choice. To characterize traffic between nodes k and l along any mode, we characterize the share of goods that are being sourced from any location i to any location j and use the link kl along the way, i.e.

$$\begin{aligned}
&\Xi_{kl} = \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}}\pi_{ij,kl}E_j \iff \\
&\Xi_{kl} = \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}}\pi_{ij}^{kl}X_{ij} \iff \\
&\Xi_{kl} = \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}}\frac{\tau_{ik}^{-\theta}t_{kl}^{-\theta}\tau_{lj}^{-\theta}}{\tau_{ij}^{-\theta}}\times\tau_{ij}^{-\theta}\frac{Y_i}{\Pi_i^{-\theta}}\frac{E_j}{P_j^{-\theta}} \iff \\
&\Xi_{kl} = t_{kl}^{-\theta}\sum_{i\in\mathcal{N}}\tau_{ik}^{-\theta}\frac{Y_i}{\Pi_i^{-\theta}}\sum_{j\in\mathcal{N}}\tau_{lj}^{-\theta}\frac{E_j}{P_j^{-\theta}}, \\
&\Xi_{kl} = t_{kl}^{-\theta}\times P_k^{-\theta}\times\Pi_l^{-\theta}
\end{aligned}$$

where in the last line we used the definition of the consumer and producer market access terms. Furthermore, replacing market access terms,

$$\begin{aligned}
P_i &= \frac{1}{\bar{W}}\bar{u}_i L_i^{\beta-1}Y_i \\
\Pi_i &= \bar{A}_i L_i^{1+\alpha}Y_i^{-\frac{\theta+1}{\theta}}
\end{aligned}$$

we obtain,

$$\begin{aligned}\Xi_{kl} &= t_{kl}^{-\theta} \times \left(\frac{1}{\bar{W}}\bar{u}_k L_k^{\beta-1} Y_k\right)^{-\theta} \times \left(\bar{A}_l L_l^{1+\alpha} Y_l^{-\frac{\theta+1}{\theta}}\right)^{-\theta} \\ &= t_{kl}^{-\theta}\left(\frac{\bar{L}^{-(\alpha+\beta)\theta}}{W^{-\theta}}\right)\bar{L}\bar{A}_l^{-\theta}\bar{u}_k^{-\theta} l_k^{-\theta(\beta-1)} l_l^{-\theta(1+\alpha)} y_k^{-\theta} y_l^{(1+\theta)} \\ &= t_{kl}^{-\theta}\chi\bar{L}\bar{A}_l^{-\theta}\bar{u}_k^{-\theta} l_k^{-\theta(\beta-1)} l_l^{-\theta(1+\alpha)} y_k^{-\theta} y_l^{(1+\theta)}\end{aligned}$$

where in the last line we use the definition $\chi = \frac{\bar{L}^{(\alpha+\beta)\theta}}{W^{\theta}}$. We have,

$$\Xi_{kl} = t_{kl}^{-\theta}\chi^{-1}\bar{L}\bar{A}_l^{-\theta}\bar{u}_k^{-\theta} l_k^{-\theta} l_l^{-\theta} y_k^{-\theta} y_l^{(1+\theta)}$$

which characterizes aggregate flows between neighboring nodes k and l in terms of the endogenous variables along the network.

A.6.2 Traffic Flows on the Mode-Specific Networks

In the next step we now turn towards characterizing the probability of sourcing from location j to location i choosing neighboring node k as the cost-minimizing routing choice and furthermore opting for mode m between node i and k. The nested choice implies that this choice probability is given by,

$$\pi_{ijk,m} = \frac{\tilde{t}_{ik,m}^{-\eta}}{t_{ik}^{-\eta}} \frac{\left(t_{ik}\tau_{kj}\right)^{-\theta} p_i^{-\theta}}{\sum_{i\in\mathcal{N}} \tau_{ij}^{-\theta} p_i^{-\theta}}$$

which can be decomposed in the sourcing share, link choice probability conditional on sourcing choice, and the modal share conditional on both sourcing and link choice, i.e.

$$\begin{aligned}\pi_{ij,kl,m} &= \frac{\tilde{t}_{ik,m}^{-\eta}}{t_{ik}^{-\eta}} \frac{\left(t_{ik}\tau_{ki}\right)^{-\theta}}{\sum_{k'\in\mathcal{F}(i)}\left(t_{ik'}\tau_{k'i}\right)^{-\theta}} \frac{\tau_{ij}^{-\theta} p_i^{-\theta}}{\sum_{i\in\mathcal{N}} \tau_{ij}^{-\theta} p_i^{-\theta}} \\ &= \pi_{ik}^{m} \times \pi_{ij}^{k} \times \pi_{ij}\end{aligned}$$

We can apply similar calculations as above to extend this for any link kl when transporting goods from i to j, i.e.

$$\pi_{ij,kl,m} = \frac{\tilde{t}_{kl,m}^{-\eta}}{t_{kl}^{-\eta}} \frac{\tau_{ik}^{-\theta} t_{kl}^{-\theta} \tau_{lj}^{-\theta}}{\tau_{ij}^{-\theta}} \frac{\tau_{ij}^{-\theta} p_i^{-\theta}}{\sum_{i\in\mathcal{N}} \tau_{ij}^{-\theta} p_i^{-\theta}}$$

Given this choice probability we can characterize the mode-specific traffic between neighboring nodes k and l, i.e.

$$\begin{aligned}\Xi_{kl,m} &= \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}} \pi_{ijk,m} E_j \iff \\ &= \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}} \pi_{ik}^{m} \times \pi_{ij}^{k} \times X_{ij} \\ &= \frac{\tilde{t}_{kl,m}^{-\eta}}{t_{kl}^{-\eta}} \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}} \frac{\tau_{ik}^{-\theta} t_{kl}^{-\theta} \tau_{lj}^{-\theta}}{\tau_{ij}^{-\theta}} \times \tau_{ij}^{-\theta} \frac{Y_i}{\Pi_i^{-\theta}} \frac{E_j}{P_j^{-\theta}} \iff \\ &= \frac{\tilde{t}_{kl,m}^{-\eta}}{t_{kl}^{-\eta}} \times t_{kl}^{-\theta} \sum_{i\in\mathcal{N}} \tau_{ik}^{-\theta} \frac{Y_i}{\Pi_i^{-\theta}} \sum_{j\in\mathcal{N}} \tau_{lj}^{-\theta} \frac{E_j}{P_j^{-\theta}}, \\ &= \frac{\tilde{t}_{kl,m}^{-\eta}}{t_{kl}^{-\eta}} \times t_{kl}^{-\theta} \times P_k^{-\theta} \times \Pi_l^{-\theta} \\ &= \tilde{t}_{kl,m}^{-\eta} \times t_{kl}^{\eta-\theta} \times P_k^{-\theta} \times \Pi_l^{-\theta}\end{aligned}$$

which gives us an expression for mode-specific traffic in terms of market access measures and the aggregate and mode specific iceberg transport cost along the edge. Note, that the nested formulation implies that mode-specific traffic is the conditional

mode specific share of aggregate traffic, i.e. $\Xi_{kl,m} = \pi^m_{ik} \times \Xi_{kl}$.

A.6.3 Traffic Flows at Terminals

In the final step we characterize the traffic at mode-specific terminals that allow multimodal movements between nodes kl. To do so we start by characterizing the probability of sourcing from location j to location i choosing any neighboring node $k' \in \mathscr{F}(i)$, but crucially choosing an alternative non-primary mode of transport and therefore traversing through a terminal while routing. This choice probability can be characterized in the following way,

$$\begin{aligned}\pi_{ij,kk,m} &= \sum_{k\in\mathscr{F}(i)} \frac{\tilde{t}^{-\eta}_{ik,m}}{t^{-\eta}_{ik}} \frac{t^{-\theta}_{ik}\tau^{-\theta}_{kj}p^{-\theta}_{j}}{\sum_{j\in 0}\sum_{k'} t_{ik}\tau_{k_jj}p^{-\theta}_{j}} \\ &= s^{-\eta}_{ii} \sum_{k\in\mathscr{F}(i)} \frac{\left(\tau_{ik}s_{kk,m}\right)^{-\eta}}{t^{-\eta}_{ik}} \frac{t^{-\theta}_{ik}\tau^{-\theta}_{kj}p^{-\theta}_{j}}{\sum_{j\in 0}\sum_{k'} t_{ik}\tau_{k_jj}p^{-\theta}_{j}}\end{aligned}$$

which can be decomposed into which can again be decomposed into sourcing shares, link choice shares conditional on sourcing choice, and mode choice conditional on link choice, i.e.

$$\begin{aligned}\pi_{ij,kk,m} &= \sum_{k\in\mathscr{F}(i)} \frac{s^{-\eta}_{ii}\left(\tau_{ik}s_{kk,m}\right)^{-\eta}}{t^{-\eta}_{ik}} \frac{\left(t_{ik}\tau_{ki}\right)^{-\theta}}{\sum_{k'\in\mathscr{F}(i)}\left(t_{ik'}\tau_{k'i}\right)^{-\theta}} \frac{\tau^{-\theta}_{ij}p^{-\theta}_{i}}{\sum_{i\in\mathscr{N}}\tau^{-\theta}_{ij}p^{-\theta}_{i}} \\ &= \pi_{ij}\sum_{k\in\mathscr{F}(i)} \pi^m_{ik} \times \pi^k_{ij}\end{aligned}$$

We can apply similar calculations as above to extend this for any link kl when transporting goods from i to j, i.e.

$$\begin{aligned}\pi_{ij,kk,m} &= \sum_{l\in\mathscr{F}(k)} \frac{s^{-\eta}_{kk}\left(\tau_{kl}s_{ll,m}\right)^{-\eta}}{t^{-\eta}_{kl}} \frac{\tau^{-\theta}_{ik}t^{-\theta}_{kl}\tau^{-\theta}_{lj}}{\tau^{-\theta}_{ij}} \frac{\tau^{-\theta}_{ij}p^{-\theta}_{i}}{\sum_{i\in\mathscr{N}}\tau^{-\theta}_{ij}p^{-\theta}_{i}} \\ &= \pi_{ij}\sum_{l\in\mathscr{F}(k)} \pi^m_{kl} \times \pi^{kl}_{ij}\end{aligned}$$

Given this choice probability we can characterize the mode-specific traffic between neighboring nodes k and l, i.e.

$$\begin{aligned}\Xi_{kk,m} &= \sum_{i\in\mathscr{N}}\sum_{j\in\mathscr{N}} \pi_{ij,kk,m}E_j \iff \\ &= \sum_{i\in\mathscr{N}}\sum_{j\in\mathscr{N}}\left(\sum_{l\in\mathscr{F}(k)} \pi^m_{kl} \times \pi^{kl}_{ij}\right) \times X_{ij} \\ &= \sum_{l\in\mathscr{F}(k)} \frac{s^{-\eta}_{kk}\left(\tau_{kl}s_{ll,m}\right)^{-\eta}}{t^{-\eta}_{kl}} \sum_{i\in\mathscr{N}}\sum_{j\in\mathscr{N}} \frac{\tau^{-\theta}_{ik}t^{-\theta}_{kl}\tau^{-\theta}_{lj}}{\tau^{-\theta}_{ij}} \times \tau^{-\theta}_{ij}\frac{Y_i}{\Pi^{-\theta}_{i}}\frac{E_j}{P^{-\theta}_{j}} \iff \\ &= \sum_{l\in\mathscr{F}(k)} \frac{s^{-\eta}_{kk}\left(\tau_{kl}s_{ll,m}\right)^{-\eta}}{t^{-\eta}_{kl}} \times t^{-\theta}_{kl}\sum_{i\in\mathscr{N}}\tau^{-\theta}_{ik}\frac{Y_i}{\Pi^{-\theta}_{i}}\sum_{j\in\mathscr{N}}\tau^{-\theta}_{lj}\frac{E_j}{P^{-\theta}_{j}}, \\ &= \sum_{l\in\mathscr{F}(k)} \frac{s^{-\eta}_{kk}\left(\tau_{kl}s_{ll,m}\right)^{-\eta}}{t^{-\eta}_{kl}} \times t^{-\theta}_{kl} \times P^{-\theta}_{k} \times \Pi^{-\theta}_{l} \\ &= s^{-\eta}_{kk} \times P^{-\theta}_{k} \times \sum_{l\in\mathscr{F}(k)}\left(\tau_{kl}s_{ll,m}\right)^{-\eta} \times t^{\eta-\theta}_{kl} \times \Pi^{-\theta}_{l}\end{aligned}$$

which gives us Equation (3.10) in Section 3.5.

A.6.4 Congestion on the Primary Transport Network

Incorporating congestion, for the primary mode of transport ($m = 1$), we have the following relationship as in the first part of in Equation (3.12) (Section 3.5),

$$
\begin{aligned}
t_{kl,1} &= \bar{t}_{kl,1}\left[\Xi_{ij,1}\right]^{\lambda_1} \iff \\
t_{kl,1} &= \bar{t}_{kl,1}\left[t_{kl,1}^{-\eta} \times t_{kl}^{\eta-\theta} \times P_k^{-\theta} \times \Pi_l^{-\theta}\right]^{\lambda_1} \iff \\
t_{kl,1} &= \bar{t}_{kl,1} \times t_{kl,1}^{-\eta\lambda_1} \times t_{kl}^{\lambda_1(\eta-\theta)} \times P_k^{-\theta\lambda_1} \times \Pi_l^{-\theta\lambda_1} \iff \\
t_{kl,1}^{1+\eta\lambda_1} &= \bar{t}_{kl,1} \times t_{kl}^{\lambda_1(\eta-\theta)} \times P_k^{-\theta\lambda_1} \times \Pi_l^{-\theta\lambda_1} \iff \\
t_{kl,1} &= \bar{t}_{kl,1}^{\frac{1}{1+\eta\lambda_1}} \times t_{kl}^{\frac{\lambda_1(\eta-\theta)}{1+\eta\lambda_1}} \times P_k^{\frac{-\theta\lambda_1}{1+\eta\lambda_1}} \times \Pi_l^{\frac{-\theta\lambda_1}{1+\eta\lambda_1}}
\end{aligned}
$$

Substituting the equation above into to derive the expression for traffic flows as in the second part of in Equation (3.12) (Section 3.5),

$$
\begin{aligned}
\Xi_{kl,1} &= \left(\bar{t}_{kl,1}^{\frac{1}{1+\theta\lambda_1}} \times t_{kl}^{\frac{\lambda_1(\eta-\theta)}{1+\eta\lambda_1}} \times P_k^{-\frac{\theta\lambda_1}{1+\theta\lambda_1}} \times \Pi_l^{-\frac{\theta\lambda_1}{1+\lambda_1}}\right)^{-\theta} \times P_k^{-\theta} \times \Pi_l^{-\theta} \\
&= \bar{t}_{kl,1}^{-\frac{\theta}{1+\theta\lambda_1}} \times t_{kl}^{\frac{-\theta\lambda_1(\eta-\theta)}{1+\eta\lambda_1}} \times P_k^{\frac{\theta\theta\lambda_1}{1+\theta\lambda_1}-\theta} \times \Pi_l^{\frac{\theta\lambda_1\theta}{1+\theta\lambda_1}-\theta} \\
&= \bar{t}_{kl,1}^{-\frac{\theta}{1+\theta\lambda_1}} \times t_{kl}^{\frac{-\theta\lambda_1(\eta-\theta)}{1+\eta\lambda_1)}} \times P_k^{-\frac{\theta}{1+\theta\lambda_1}} \times \Pi_l^{-\frac{\theta}{1+\theta\lambda_1}}
\end{aligned}
$$

A.6.5 Congestion at Intermodal Terminals

For any secondary mode of transportation ($m \neq 1$), we incorporate congestion at terminal stations that connect the primary and secondary networks. Let us characterize the transport cost net of congestion at any terminal, i.e.

$$
\begin{aligned}
s_{kk,m} &= \bar{s}_{kk,m}\left[\Xi_{kk,m}\right]^{\lambda_m} \\
s_{kk,m} &= \bar{s}_{kk,m}\left[s_{kk,m}^{-\eta} \times P_k^{-\theta} \times \sum_{l\in\mathscr{F}(k)}\left(\tau_{kl}s_{ll,m}\right)^{-\eta} \times t_{kl}^{\eta-\theta} \times \Pi_l^{-\theta}\right]^{\lambda_m} \\
s_{kk,m} &= \bar{s}_{kk,m} \times s_{kk,m}^{-\eta\lambda_m} \times P_k^{-\theta\lambda_m} \times \left(\sum_{l\in\mathscr{F}(k)}\left(\tau_{kl}s_{ll,m}\right)^{-\eta} \times t_{kl}^{\eta-\theta} \times \Pi_l^{-\theta}\right)^{\lambda_m} \\
s_{kk,m}^{1+\eta\lambda_m} &= \bar{s}_{kk,m} \times P_k^{-\theta\lambda_m} \times \left(\sum_{l\in\mathscr{F}(k)}\left(\tau_{kl}s_{ll,m}\right)^{-\eta} \times t_{kl}^{\eta-\theta} \times \Pi_l^{-\theta}\right)^{\lambda_m} \\
s_{kk,m} &= \bar{s}_{kk,m}^{1+n\lambda_m} \times P_k^{\frac{-\theta\lambda_m}{1+\eta\lambda_m}} \times \left(\sum_{l\in\mathscr{F}(k)}\left(\tau_{kl}s_{ll,m}\right)^{-\eta} \times t_{kl}^{\eta-\theta} \times \Pi_l^{-\theta}\right)^{\frac{\lambda_m}{1+\eta\lambda_m}} \\
s_{kk,m} &= \bar{s}_{kk,m}^{\frac{1}{1+\eta\lambda m}} \times P_k^{\frac{-\theta\lambda_m}{1+\eta\lambda_m}} \times \left(\sum_{l\in\mathscr{F}(k)}\left(\tau_{kl}s_{ll,m}\right)^{-\eta} \times t_{kl}^{\eta-\theta} \times \Pi_l^{-\theta}\right)^{\frac{\lambda_m}{1+\eta\lambda_m}}
\end{aligned}
$$

Combining both the switching cost of access the non-primary transport network and the terminal cost of exiting it again, we

obtain Equation (3.14) in Section 3.5:

$$
\begin{aligned}
t_{kl,m} &= \bar{s}_{kk,m}\tau_{kl,m}\bar{s}_{ll,m}\left[\Xi_{kk,m}\right]^{\lambda_m}\left[\Xi_{ll,m}\right]^{\lambda_m} \\
&= \bar{s}_{kk,m}\tau_{kl,m}\bar{s}_{ll,m}\left[\sum_{l\in\mathscr{F}(k)}\Xi_{kl,m}\right]^{\lambda_m}\left[\sum_{k\in\mathscr{B}(l)}\Xi_{kl,m}\right]^{\lambda_m} \\
&= \bar{s}_{kk,m}\tau_{kl,m}\bar{s}_{ll,m}\left[P_k^{-\theta}\times\sum_{l\in\mathscr{F}(k)}\tilde{t}_{kl,m}^{-\eta}\times t_{kl}^{\eta-\theta}\times\Pi_l^{-\theta}\right]^{\lambda_m}\left[\Pi_l^{-\theta}\times\sum_{k\in\mathscr{B}(l)}\tilde{t}_{kl,m}^{-\eta}\times t_{kl}^{\eta-\theta}\times P_k^{-\theta}\right]^{\lambda_m} \\
&= \bar{s}_{kk,m}\tau_{kl,m}\bar{s}_{ll,m}\left[P_k^{-\theta}\times\Pi_{k,m}^{-\theta}\right]^{\lambda_m}\left[\Pi_l^{-\theta}\times P_{l,m}^{-\theta}\right]^{\lambda_m} \\
&= \bar{s}_{kk,m}\times\tau_{kl,m}\times\bar{s}_{ll,m}\times P_k^{-\theta\lambda_m}\times\Pi_l^{-\theta\lambda_m}\times P_{k,m}^{-\theta\lambda_m}\times\Pi_{l,m}^{-\theta\lambda_m}
\end{aligned}
$$

where $P_{l,m}\equiv\sum_{k\in\mathscr{B}(l)}t_{kl,m}^{-\eta}\times t_{kl}^{\eta-\theta}\times P_k^{-\theta}$ and $\Pi_{k,m}\equiv\sum_{l\in\mathscr{F}(k)}t_{kl,m}^{-\eta}\times t_{kl}^{\eta-\theta}\times\Pi_l^{-\theta}$.

A.7 Derivations for Section 3.6: Traffic Substitution and Complementarity

This subsection develops the traffic-adjustment view used at the beginning of Section 3.6. The aim is to show how a mode-specific change in transport cost on link kl affects mode-specific traffic on that same link. To keep the mechanism transparent, the derivation abstracts here from congestion and focuses on the recursive route block with nested mode choice.

Start from

$$\Xi_{kl,m} = \pi^m_{kl}\,\Xi_{kl}, \qquad \pi^m_{kl} = \frac{t_{kl,m}^{-\eta}}{t_{kl}^{-\eta}}, \qquad \Xi_{kl} = t_{kl}^{-\theta} P_k^{-\theta} \Pi_l^{-\theta}, \tag{A.2}$$

with

$$t_{kl} = \Big(\sum_{n \in M(k,l)} t_{kl,n}^{-\eta} \Big)^{-1/\eta}, \qquad \varpi_{kl,m} := \frac{\Xi_{kl,m}}{\sum_n \Xi_{kl,n}} = \pi^m_{kl}. \tag{A.3}$$

Define the transport-propagation matrices

$$\Omega^{t,P} = \Big[\frac{t_{kl}^{-\theta} P_k^{-\theta}}{P_l^{-\theta}} \Big]_{kl}, \qquad \Omega^{t,\Pi} = \Big[\frac{t_{kl}^{-\theta} \Pi_l^{-\theta}}{\Pi_k^{-\theta}} \Big]_{kl}, \tag{A.4}$$

and their Leontief-type transforms

$$\widetilde{\Omega}^{t,P} := (I - \Omega^{t,P})^{-1} \Omega^{t,P}, \qquad \widetilde{\Omega}^{t,\Pi} := (I - \Omega^{t,\Pi})^{-1} \Omega^{t,\Pi}. \tag{A.5}$$

Proposition 2 (Traffic substitution and complementarity)**.** *For an alternative mode $m \neq m'$ on link kl,*

$$\frac{d \ln \Xi_{kl,m}}{d \ln t_{kl,m'}} = \epsilon^{\pi^m_{kl}}_{t_{kl,m'}} + \epsilon^{\Xi_{kl}}_{t_{kl}} \epsilon^{t_{kl}}_{t_{kl,m'}} \approx \Big[\eta - \theta \big(1 + [\widetilde{\Omega}^{t,P}]_{kl} + [\widetilde{\Omega}^{t,\Pi}]_{kl} \big) \Big] \varpi_{kl,m'}. \tag{A.6}$$

Moreover,

$$\epsilon^{\pi^m_{kl}}_{t_{kl,m}} = -\eta(1 - \varpi_{kl,m}), \qquad \epsilon^{\pi^m_{kl}}_{t_{kl,m'}} = \eta \varpi_{kl,m'} \quad (m \neq m'), \qquad \epsilon^{t_{kl}}_{t_{kl,m'}} = \varpi_{kl,m'}. \tag{A.7}$$

Proof. Totally differentiating (A.2) gives

$$\frac{d \ln \Xi_{kl,m}}{d \ln t_{kl,m'}} = \epsilon^{\pi^m_{kl}}_{t_{kl,m'}} + \epsilon^{\Xi_{kl}}_{t_{kl}} \epsilon^{t_{kl}}_{t_{kl,m'}}.$$

Differentiating the conditional mode share and the aggregate link cost index yields (A.7). Differentiating aggregate traffic gives

$$\epsilon^{\Xi_{kl}}_{t_{kl}} = -\theta + \epsilon^{P_k^{-\theta}}_{t_{kl}} + \epsilon^{\Pi_l^{-\theta}}_{t_{kl}}. \tag{A.8}$$

Abstracting from changes in local expenditure and sales shifters, the recursive market-access system implies the approximations

$$\epsilon^{P_k^{-\theta}}_{t_{kl}} \approx -\theta [\widetilde{\Omega}^{t,P}]_{kl}, \qquad \epsilon^{\Pi_l^{-\theta}}_{t_{kl}} \approx -\theta [\widetilde{\Omega}^{t,\Pi}]_{kl}. \tag{A.9}$$

Substituting (A.7) and (A.9) into (A.8) yields (A.6). □

Proposition 2 is the traffic counterpart of the welfare wedges used in the main text. The first term is direct substitution across modes on the same link. The second is indirect complementarity: changing the cost of one mode changes the aggregate cost of the link, which changes total traffic and market access across the network. Central links therefore generate stronger complementarity effects because improving them changes more than the local modal split.

A.8 Derivations for Section 3.6: Welfare Elasticities, Wedges, and Signs

The previous subsection described how a mode-specific cost change redistributes traffic through direct substitution and indirect complementarity. We now turn to the welfare-elasticity view and ask how a lower mode-specific cost on link kl translates into welfare under different model variants. To do so, we develop a set of decompositional results that build on the inverse function theorem applied to this class of models as previously developed in Allen, Fuchs and Wong (2025). The qualitative connection is the same as in the main text: the welfare formulas value the same route and mode reallocation, congestion across links and at terminals, and market-access propagation emphasized in the traffic block. For each model variant c, let the transformed equilibrium satisfy $G_c(\zeta) = 0$ and write $J_c = \partial G_c / \partial \zeta$ for the Jacobian of that system. Applying the inverse function theorem gives the welfare effect of a marginal improvement in mode m on link kl as

$$E^c_{kl,m} = -g^\top J_c^{-1} b_{kl,m}. \tag{A.10}$$

The purpose of this appendix is to turn (A.11) into the closed-form objects discussed in Section 3.6. Proposition 5 fully characterizes the road and terminal wedges in closed form as the load placed by a mode-specific link improvement on congested road or terminal states times the welfare value of relaxing those states. Proposition 6, the main result for the reallocation side, fully characterizes the mode and route wedges in closed form as local reallocation, induced congestion, and broader equilibrium feedback. The final two subsections record the qualitative sign results used in the main text and the compact transformed-equilibrium ingredients behind the formulas.

A.8.1 Sparse Jacobian representation and exact ladder.

Throughout the transport-side comparisons, the same initial mode-specific link improvement is evaluated at the observed transport assignment, so the welfare row and shock map are held fixed across model variants. For each model variant c, let the transformed equilibrium satisfy $G_c(\zeta) = 0$ and write $J_c = \partial G_c / \partial \zeta$ for the Jacobian of that system. The inverse function theorem gives the welfare effect of a marginal improvement in mode m on link kl as

$$E^c_{kl,m} = -g^\top J_c^{-1} b_{kl,m}. \tag{A.11}$$

Here $g^\top$ is the welfare row that maps changes in the transformed equilibrium variables into changes in $\ln W$, and $b_{kl,m}$ is the perturbation induced in the transformed equilibrium equations by the marginal reduction in the primitive cost of mode m on link kl. Writing $\rho := 1 + \alpha + \beta$, let B_t denote the common-baseline version of the cost-loading operator that maps composite-link cost changes into the transformed equilibrium equations, and define the composite-link shock

$$b_{kl} := B_t \mathbf{e}_{kl}, \qquad s_{kl} := b_{kl}/\Xi_{kl}, \qquad \omega^\top := -g^\top/\rho, \qquad q_c^\top := \omega^\top J_c^{-1}, \qquad r^c_{kl} := J_c^{-1} s_{kl}. \tag{A.12}$$

Here $\mathbf{e}_{kl}$ is the basis vector for a marginal reduction in the composite cost of link kl. The row vector $q_c^\top$ is the welfare adjoint under model variant c, and the column vector r^c_{kl} is the equilibrium response to the normalized composite-link shock s_{kl}. A primitive mode-specific improvement enters through the nested mode aggregator. Define

$$\vartheta_{kl,m} := \left.\frac{\partial \ln t}{\partial \ln \bar{t}_{kl,m}}\right|_\zeta, \qquad b_{kl,m} := B_t \vartheta_{kl,m}, \qquad s_{kl,m} := b_{kl,m}/\Xi_{kl,m}. \tag{A.13}$$

Under the nested CES link-cost index,

$$\vartheta_{kl,m} = \pi^m_{kl} \mathbf{e}_{kl}, \qquad b_{kl,m} = \pi^m_{kl} b_{kl}, \qquad s_{kl,m} = s_{kl}, \tag{A.14}$$

where the last equality uses $\Xi_{kl,m} = \pi^m_{kl} \Xi_{kl}$.

Proposition 3 (Normalized welfare object)**.** *For each model variant c,*

$$m^c_{kl} = \omega^\top J_c^{-1} s_{kl} = \omega^\top J_c^{-1} s_{kl,m} = q_c^\top s_{kl} = \omega^\top r^c_{kl}, \tag{A.15}$$

and the mode-specific welfare elasticity satisfies

$$E^c_{kl,m} = \rho\, \Xi_{kl,m}\, m^c_{kl} = \pi^m_{kl} \rho\, \Xi_{kl}\, m^c_{kl}. \tag{A.16}$$

In particular, the normalized multiplier is link-specific in the reduced equilibrium representation, while the mode label enters through the exposure term $\Xi_{kl,m}$.

Proof. Substitute $b_{kl,m} = \Xi_{kl,m} s_{kl,m}$ into (A.11), use (A.14), and divide by ρ. □

To keep the appendix aligned with Section 3.6, we organize the welfare formulas directly around the Jacobian blocks that generate the ladder. Define

$$J_c = R_c + \Delta_{\text{road},c} + \Delta_{\text{term},c} + G_c, \qquad G_c := u_c v_c^\top, \tag{A.17}$$

where R_c is the grouped sparse reallocation block of the Jacobian, collecting the local own-share and sparse forward/backward propagation terms; $\Delta_{\text{road},c}$ and $\Delta_{\text{term},c}$ are the road- and terminal-congestion feedback blocks; and G_c is the low-rank aggregate block, with u_c and v_c the vectors entering the aggregate component of the Jacobian. Appendix A.8.5 derives all four blocks explicitly.

A useful way to read the ladder is as an *adjoint–block–response* decomposition. For any comparison (A, B), the response vector $r_{kl}^B = J_B^{-1} s_{kl}$ records the footprint of the link improvement on the equilibrium system under model variant B, the Jacobian difference $J_B - J_A$ records which propagation coefficients change when a given mechanism is turned on or off, and the adjoint vector $q_A^\top = \omega^\top J_A^{-1}$ weights those affected equations by their welfare shadow value under model variant A. Under the common-baseline convention, all of these objects are evaluated around the same observed transport assignment. The ladder therefore does not compare different baselines; it is a local comparative-static anatomy of the same baseline equilibrium, decomposed by the Jacobian blocks that govern sparse reallocation, congestion feedback, and aggregate market-access adjustment.

Economically, each block can be read as a distinct adjustment margin that is switched on or off around the same baseline allocation. The sparse reallocation block R_c is the local choice-set margin: it governs whether, after the link improves, traffic can exploit additional route substitutions, additional mode substitutions, and additional feeder or sourcing continuations around that corridor. The road and terminal blocks are the scarcity margins: they determine how much of that extra local adjustment is taxed away once the induced traffic is allowed to congest links and terminals. The low-rank block G_c is the broad market-access margin: it determines whether the local transport gain is amplified or diluted once activity and demand are allowed to rebalance through the common equilibrium channels. When two model variants are close in Jacobian space, so that $J_B - J_A$ is small, we may replace r_{kl}^B by r_{kl}^A and obtain the first-order reading

$$d_{kl}(A, B) \approx q_A^\top (J_B - J_A) r_{kl}^A. \tag{A.18}$$

This approximation says that a wedge is the welfare value of passing the baseline footprint of the improved link through the additional propagation coefficients associated with the mechanism being turned on or off.

Proposition 4 (Exact ladder from Jacobian blocks)**.** *For any two model variants A, B,*

$$d_{kl}(A, B) := m_{kl}^A - m_{kl}^B = q_A^\top (J_B - J_A) r_{kl}^B. \tag{A.19}$$

Using (A.17)*, the gap admits the block decomposition*

$$d_{kl}(A, B) = d_{kl}^{\text{alloc}}(A, B) + d_{kl}^{\text{road}}(A, B) + d_{kl}^{\text{term}}(A, B) + d_{kl}^{\text{glob}}(A, B), \tag{A.20}$$

where

$$d_{kl}^{\text{alloc}}(A, B) := q_A^\top (R_B - R_A) r_{kl}^B, \tag{A.21}$$

$$d_{kl}^{\text{road}}(A, B) := q_A^\top (\Delta_{\text{road},B} - \Delta_{\text{road},A}) r_{kl}^B, \tag{A.22}$$

$$d_{kl}^{\text{term}}(A, B) := q_A^\top (\Delta_{\text{term},B} - \Delta_{\text{term},A}) r_{kl}^B, \tag{A.23}$$

$$d_{kl}^{\text{glob}}(A, B) := q_A^\top (G_B - G_A) r_{kl}^B. \tag{A.24}$$

Equivalently, letting

$$d_{kl}^{\text{cong}}(A, B) := d_{kl}^{\text{road}}(A, B) + d_{kl}^{\text{term}}(A, B), \tag{A.25}$$

we obtain the coarse decomposition used in Section 3.6:

$$d_{kl}(A, B) = d_{kl}^{\text{alloc}}(A, B) + d_{kl}^{\text{cong}}(A, B) + d_{kl}^{\text{glob}}(A, B). \tag{A.26}$$

In particular,

$$d_{kl}^{\text{road}} = d_{kl}(\text{NC}, \text{NT}), \qquad d_{kl}^{\text{term}} = d_{kl}(\text{NT}, F), \qquad d_{kl}^{\text{mode}} = d_{kl}(F, \text{RO}), \qquad d_{kl}^{\text{route}} = d_{kl}(F, \text{FR}). \tag{A.27}$$

Hence

$$m_{kl}^{F} = m_{kl}^{\mathrm{NC}} - d_{kl}^{\mathrm{road}} - d_{kl}^{\mathrm{term}} = m_{kl}^{\mathrm{RO}} + d_{kl}^{\mathrm{mode}} = m_{kl}^{\mathrm{FR}} + d_{kl}^{\mathrm{route}}. \tag{A.28}$$

The Hulten comparison satisfies

$$h_{kl} = \rho\big(m_{kl}^{H} - m_{kl}^{F}\big) = (1-\rho) + \rho(1 - m_{kl}^{F}), \tag{A.29}$$

Proof. The gap identity is

$$m_{kl}^{A} - m_{kl}^{B} = \omega^{\top}\big(J_A^{-1} - J_B^{-1}\big)s_{kl} = q_A^{\top}(J_B - J_A)r_{kl}^{B}.$$

Substituting (A.17) yields (A.20); grouping the two congestion blocks gives (A.26). The named wedges are the corresponding special cases in (A.27), and (A.29) follows from the efficient-equilibrium benchmark $E_{kl,m}^{H} = \Xi_{kl,m}$, equivalently $m_{kl}^{H} = 1/\rho$, together with Proposition 3. □

Proposition 4 is the main organizing object for the rest of the appendix. The pure congestion wedges isolate single congestion blocks. The mode and route wedges compare the full model to restricted variants that alter the sparse reallocation block R_c, the induced congestion blocks, and the low-rank aggregate block. This is exactly the qualitative anatomy used in Section 3.6.

A.8.2 Congestion wedges in closed form.

The road and terminal wedges are exact additive updates because moving from NC to NT restores only the road-scarcity block, while moving from NT to F restores only the terminal-scarcity block. Recall that these congestion blocks admit the factorization

$$\Delta_{\mathrm{road},c} = B_{t,c}K_{\mathrm{road},c}Q_{\mathrm{road},c}, \qquad \Delta_{\mathrm{term},c} = B_{t,c}K_{\mathrm{term},c}Q_{\mathrm{term},c},$$

where $B_{t,c} := \partial G_c/\partial \ln t^c$ loads composite transport-cost changes into the transformed equilibrium equations, $Q_{\mathrm{road},c} := \partial \ln \Xi^c/\partial \zeta^{\top}$ and $Q_{\mathrm{term},c} := \partial \ln T^c/\partial \zeta^{\top}$ map equilibrium responses into road-congestion states Ξ^c and terminal-congestion states T^c, and $K_{\mathrm{road},c}$ and $K_{\mathrm{term},c}$ map those congestion-state changes back into link-cost changes. In particular,

$$J_{\mathrm{NT}} - J_{\mathrm{NC}} = \Delta_{\mathrm{road,NT}}, \qquad J_F - J_{\mathrm{NT}} = \Delta_{\mathrm{term},F}. \tag{A.30}$$

Proposition 5 (Exact congestion wedges)**.** *Define*

$$\begin{aligned} v_{\mathrm{road}}^{\mathrm{NC:NT}\top} &:= q_{\mathrm{NC}}^{\top}B_{t,\mathrm{NT}}K_{\mathrm{road,NT}}, & \ell_{kl,\mathrm{road}}^{\mathrm{NT}} &:= Q_{\mathrm{road,NT}}r_{kl}^{\mathrm{NT}}, \\ v_{\mathrm{term}}^{\mathrm{NT:}F\top} &:= q_{\mathrm{NT}}^{\top}B_{t,F}K_{\mathrm{term},F}, & \ell_{kl,\mathrm{term}}^{F} &:= Q_{\mathrm{term},F}r_{kl}^{F}. \end{aligned} \tag{A.31}$$

Then

$$d_{kl}^{\mathrm{road}} = q_{\mathrm{NC}}^{\top}\Delta_{\mathrm{road,NT}}r_{kl}^{\mathrm{NT}} = v_{\mathrm{road}}^{\mathrm{NC:NT}\top}\ell_{kl,\mathrm{road}}^{\mathrm{NT}}, \qquad d_{kl}^{\mathrm{term}} = q_{\mathrm{NT}}^{\top}\Delta_{\mathrm{term},F}r_{kl}^{F} = v_{\mathrm{term}}^{\mathrm{NT:}F\top}\ell_{kl,\mathrm{term}}^{F}. \tag{A.32}$$

By construction,

$$d_{kl}^{\mathrm{scar}} = d_{kl}^{\mathrm{road}} + d_{kl}^{\mathrm{term}}. \tag{A.33}$$

Proof. Specialize Proposition 4 to the two updates in (A.30) and substitute the factorization in (A.73). □

Proposition 5 shows that the pure congestion wedges are the interaction between two objects: the load that the mode-specific link improvement places on congested road or terminal states, and the welfare value of relaxing those states. This is the formal counterpart of the congestion-wedge interpretation in Section 3.6.

A.8.3 Reallocation wedges in closed form.

The mode and route wedges compare the full model to restricted variants that suppress either non-road modal diversion or endogenous rerouting. Relative to the full model, these comparisons change the sparse reallocation block R_c, the induced congestion blocks, and the low-rank aggregate block. Concretely, $R_A - R_F$ records which sparse local reallocation coefficients are missing in the restricted variant $A \in \{\mathrm{RO}, \mathrm{FR}\}$, $(\Delta_{\mathrm{road},A} - \Delta_{\mathrm{road},F}) + (\Delta_{\mathrm{term},A} - \Delta_{\mathrm{term},F})$ records the associated change in scarcity feedback, and $G_A - G_F$ records the corresponding change in the low-rank aggregate feedback block.

Proposition 6 (Exact reallocation wedges). *For $A \in \{\mathrm{RO}, \mathrm{FR}\}$, define*

$$d_{kl}^{\mathrm{alloc}}(A) := d_{kl}^{\mathrm{alloc}}(F,A) = q_F^\top (R_A - R_F) r_{kl}^A, \tag{A.34}$$

$$d_{kl}^{\mathrm{cong}}(A) := d_{kl}^{\mathrm{cong}}(F,A) = q_F^\top \Big[(\Delta_{\mathrm{road},A} - \Delta_{\mathrm{road},F}) + (\Delta_{\mathrm{term},A} - \Delta_{\mathrm{term},F}) \Big] r_{kl}^A, \tag{A.35}$$

$$d_{kl}^{\mathrm{glob}}(A) := d_{kl}^{\mathrm{glob}}(F,A) = q_F^\top (G_A - G_F) r_{kl}^A. \tag{A.36}$$

Then

$$d_{kl}(F,A) = d_{kl}^{\mathrm{alloc}}(A) + d_{kl}^{\mathrm{cong}}(A) + d_{kl}^{\mathrm{glob}}(A). \tag{A.37}$$

In particular,

$$d_{kl}^{\mathrm{mode}} = d_{kl}(F, \mathrm{RO}), \qquad d_{kl}^{\mathrm{route}} = d_{kl}(F, \mathrm{FR}). \tag{A.38}$$

If

$$\begin{aligned} &d_{kl,\mathrm{alloc}}^{\mathrm{mode}} := d_{kl}^{\mathrm{alloc}}(\mathrm{RO}), && d_{kl,\mathrm{cong}}^{\mathrm{mode}} := d_{kl}^{\mathrm{cong}}(\mathrm{RO}), && d_{kl,\mathrm{glob}}^{\mathrm{mode}} := d_{kl}^{\mathrm{glob}}(\mathrm{RO}), \\ &d_{kl,\mathrm{alloc}}^{\mathrm{route}} := d_{kl}^{\mathrm{alloc}}(\mathrm{FR}), && d_{kl,\mathrm{cong}}^{\mathrm{route}} := d_{kl}^{\mathrm{cong}}(\mathrm{FR}), && d_{kl,\mathrm{glob}}^{\mathrm{route}} := d_{kl}^{\mathrm{glob}}(\mathrm{FR}), \end{aligned} \tag{A.39}$$

then

$$d_{kl}^{\mathrm{mode}} = d_{kl,\mathrm{alloc}}^{\mathrm{mode}} + d_{kl,\mathrm{cong}}^{\mathrm{mode}} + d_{kl,\mathrm{glob}}^{\mathrm{mode}}, \qquad d_{kl}^{\mathrm{route}} = d_{kl,\mathrm{alloc}}^{\mathrm{route}} + d_{kl,\mathrm{cong}}^{\mathrm{route}} + d_{kl,\mathrm{glob}}^{\mathrm{route}}. \tag{A.40}$$

Proof. Equation (A.37) is the specialization of (A.26) to $(A,B) = (F,A)$. □

Proposition 6 matches the qualitative language in Section 3.6. The term $d_{kl}^{\mathrm{alloc}}(A)$ is the exact welfare contribution of replacing the restricted model's sparse reallocation block by the full-model block. Economically, this is the value of giving the economy a richer locally relevant choice set around the improved link: additional route substitutions, additional mode substitutions, and additional feeder or sourcing continuations. If $d_{kl}^{\mathrm{alloc}}(A) > 0$, those extra options are complements to the improved corridor, because restoring them makes the link more useful within the network. If $d_{kl}^{\mathrm{alloc}}(A) < 0$, the extra options mainly act as substitutes or bypasses, because restoring them makes the improved link easier to avoid.

The term $d_{kl}^{\mathrm{cong}}(A)$ is the scarcity penalty on that extra flexibility. Once those additional local continuations are available, the network may exploit the improved link more aggressively, but the resulting traffic can spill onto congested roads and congested terminals. A positive induced-congestion term means that restoring the missing flexibility also restores the scarcity feedback that absorbs part of the gain. A negative induced-congestion term would mean that the extra flexibility helps the network steer traffic away from bottlenecks rather than into them.

The term $d_{kl}^{\mathrm{glob}}(A)$ records broader equilibrium feedback through the low-rank aggregate block. This is the system-wide rebalancing margin: once the local transport improvement changes access, activity and demand can move through a small number of common market-access channels. A positive aggregate term means that this broader rebalancing amplifies the local gain; a negative aggregate term means that it dilutes or offsets it.

A useful first-order reading is obtained by replacing r_{kl}^A by r_{kl}^F. Then the comparison to a restricted variant can be read as taking the baseline full-model propagation of the link improvement and passing it through the Jacobian blocks that are shut down in that restricted model, weighted by the welfare shadow values in the full model. In this first-order sense, the ladder asks whether the marginal incidence of the improved link becomes more valuable when the economy is allowed to use more routes and modes, or whether those extra options mainly interact with congestion and broader equilibrium feedback to reduce the pass-through to welfare.

A.8.4 Qualitative sign results.

The exact formulas above do not require sign restrictions. The next result records the one-sided sign statements that support the economic interpretation in Section 3.6: the pure congestion wedges are positive under one-sided propagation conditions, the induced-congestion piece is positive when reallocation relieves scarcity in the right places, the broader-feedback term is bounded but not sign-definite, and a positive full reallocation wedge requires local gains to dominate that broader feedback.

Corollary 1 (Qualitative sign results and dominance criterion). *Assume sign-aligned propagation on the active support: whenever the kernels below are entrywise nonnegative, the relevant adjoint and response vectors lie in the same closed orthant when paired with them.*

1. *If $\Delta_{\mathrm{road,NT}} \geq 0$ and $\Delta_{\mathrm{term},F} \geq 0$ entrywise on the active support, then*

$$d_{kl}^{\mathrm{road}} \geq 0, \qquad d_{kl}^{\mathrm{term}} \geq 0. \tag{A.41}$$

2. *Fix $A \in \{\mathrm{RO}, \mathrm{FR}\}$. If*

$$(\Delta_{\mathrm{road},A} - \Delta_{\mathrm{road},F}) + (\Delta_{\mathrm{term},A} - \Delta_{\mathrm{term},F}) \geq 0 \tag{A.42}$$

entrywise on the active support, then

$$d_{kl}^{\mathrm{cong}}(A) \geq 0. \tag{A.43}$$

3. *The broader-feedback term has no universal sign. More generally,*

$$\left| d_{kl}^{\mathrm{glob}}(A) \right| \leq \|q_F\| \, \|G_A - G_F\| \, \|r_{kl}^A\|. \tag{A.44}$$

Because $G_A - G_F$ has rank at most two, this term is low-dimensional even when its sign is ambiguous.

4. *If*

$$d_{kl}^{\mathrm{alloc}}(A) + d_{kl}^{\mathrm{cong}}(A) > \left| d_{kl}^{\mathrm{glob}}(A) \right|, \tag{A.45}$$

then

$$d_{kl}(F,A) > 0. \tag{A.46}$$

Proof. Items 1 and 2 follow directly from Proposition 5 and Proposition 6 under sign-aligned propagation. Item 3 is the standard operator-norm bound applied to (A.36). Item 4 is immediate from (A.37). □

These sign results make precise the main-text interpretation. Congestion wedges are positive when allowing congestion across links and at terminals to respond blocks exactly the margins through which a mode-specific link improvement would otherwise raise traffic and welfare. For the reallocation wedges, the induced-congestion piece is positive when local reorganization relieves congestion on links or terminals that matter a lot for the network, while the broader-feedback term may either reinforce or dilute the local gain. A positive full reallocation wedge therefore requires local reallocation gains and congestion relief to dominate that broader equilibrium feedback.

A.8.5 Deriving the sparse Jacobian representation.

For completeness, this subsection derives the transformed equilibrium system and sparse Jacobian decomposition used in Propositions 3–6. For any model variant that preserves the recursive spatial structure, equilibrium can be written as

$$A_i^{1-\sigma} w_i^{\sigma} L_i = W^{1-\sigma} u_i^{\sigma-1} w_i^{\sigma} L_i + \sum_{k \in N(i)} t_{ik}^{1-\sigma} A_k^{1-\sigma} w_k^{\sigma} L_k, \tag{A.47}$$

$$w_i^{1-\sigma} u_i^{1-\sigma} = W^{1-\sigma} A_i^{\sigma-1} w_i^{1-\sigma} + \sum_{k \in N(i)} t_{ki}^{1-\sigma} w_k^{1-\sigma} u_k^{1-\sigma}, \tag{A.48}$$

where $N(i)$ denotes the set of locations directly connected to location i. We take aggregate population $\bar{L} := \sum_i L_i$ as fixed. With

$$A_i = L_i^{\alpha}, \qquad u_i = L_i^{\beta}, \qquad \rho := 1 + \alpha + \beta. \tag{A.49}$$

Define the transformed variables

$$\varphi := 1 + \beta(\sigma - 1) + \alpha\sigma, \qquad x_i := \left(W^{-1/\rho} L_i \right)^{1+\beta(\sigma-1)} w_i^{\sigma}, \qquad y_i := \left(W^{-1/\rho} L_i \right)^{\alpha(\sigma-1)} w_i^{1-\sigma}. \tag{A.50}$$

If $\varphi \neq 0$, this map is invertible with

$$w_i = x_i^{\alpha/\varphi} y_i^{-\frac{1+\beta(\sigma-1)}{(\sigma-1)\varphi}}, \qquad L_i = W^{1/\rho} x_i^{1/\varphi} y_i^{\sigma/((\sigma-1)\varphi)}. \tag{A.51}$$

Write

$$a := \frac{1+\alpha}{\varphi}, \qquad b := \frac{\sigma(\alpha+\beta)}{\varphi}, \qquad \gamma := \frac{(\alpha+\beta)(\sigma-1)}{\varphi}, \qquad \varepsilon := 1 - b, \tag{A.52}$$

$$z_i := x_i^{1/\varphi} y_i^{\sigma/((\sigma-1)\varphi)}, \qquad \Omega := \left(\frac{1}{\bar{L}} \sum_n z_n \right)^{\sigma-1}, \tag{A.53}$$

$$R_i^x := \Omega x_i + \sum_{k \in N(i)} t_{ik}^{1-\sigma} x_k^{a} y_k^{-b}, \qquad R_i^y := \Omega y_i + \sum_{k \in N(i)} t_{ki}^{1-\sigma} x_k^{-\gamma} y_k^{\varepsilon}. \tag{A.54}$$

Then equilibrium is $G(\zeta) = 0$, with

$$G_{1i} \coloneqq \ln R_i^x - a \ln x_i + b \ln y_i = 0, \qquad i = 1, \dots, N, \tag{A.55}$$

$$G_{2i} \coloneqq \ln R_i^y + \gamma \ln x_i - \varepsilon \ln y_i = 0, \qquad i = 1, \dots, N-1, \tag{A.56}$$

$$G_{2N} \coloneqq \frac{\alpha}{\varphi} \ln x_N - \frac{1 + \beta(\sigma - 1)}{(\sigma - 1)\varphi} \ln y_N = 0. \tag{A.57}$$

Let

$$\zeta \coloneqq (\ln x_1, \dots, \ln x_N, \ln y_1, \dots, \ln y_N)^\top. \tag{A.58}$$

Define the transformed balance shares and recursive propagation shares

$$s_i^x \coloneqq \frac{\Omega x_i}{R_i^x}, \qquad \mu_{ij} \coloneqq \frac{t_{ij}^{1-\sigma} x_j^a y_j^{-b}}{R_i^x}, \tag{A.59}$$

$$s_i^y \coloneqq \frac{\Omega y_i}{R_i^y}, \qquad \lambda_{ji} \coloneqq \frac{t_{ji}^{1-\sigma} x_j^{-\gamma} y_j^{\varepsilon}}{R_i^y}. \tag{A.60}$$

If $M_c = (\mu_{ij}^c)$, $\Lambda_c = (\lambda_{ji}^c)$, $S_x^c = \operatorname{diag}(s_i^{x,c})$, $S_y^c = \operatorname{diag}(s_i^{y,c})$, and $\ell_c = (L_i^c / \bar{L})_i$, then differentiation yields

$$J_c = B_c + u_c v_c^\top + \Delta_{\text{road},c} + \Delta_{\text{term},c}, \tag{A.61}$$

with

$$B_c = \begin{bmatrix} S_x^c - aI + aM_c & bI - bM_c \\ \gamma(I - \Lambda_c^\top)_{1:(N-1),:} & (S_y^c)^{(N-1)} - \varepsilon(I - \Lambda_c^\top)_{1:(N-1),:} \\ \frac{\alpha}{\varphi} e_N^\top & -\frac{1+\beta(\sigma-1)}{(\sigma-1)\varphi} e_N^\top \end{bmatrix},$$

$$u_c = \begin{bmatrix} \frac{\sigma-1}{\varphi} s_x^c \\ \frac{\sigma-1}{\varphi} s_{y,1:(N-1)}^c \\ 0 \end{bmatrix},$$

$$v_c^\top = \begin{bmatrix} \ell_c^\top & \frac{\sigma}{\sigma-1} \ell_c^\top \end{bmatrix}. \tag{A.62}$$

Here e_N is the Nth standard basis vector, $s_x^c = (s_i^{x,c})_{i=1}^N$ and $s_{y,1:(N-1)}^c = (s_i^{y,c})_{i=1}^{N-1}$ collect the transformed balance shares, and $\ell_c = (L_i^c / \bar{L})_{i=1}^N$ collects population shares. The local block B_c governs sparse reallocation through neighboring links, $u_c v_c^\top$ is the common aggregate-propagation block, and the two Δ matrices are the additive congestion blocks. Thus the grouped sparse reallocation block used in the main line of the appendix is simply

$$R_c \coloneqq B_c, \qquad J_c = R_c + \Delta_{\text{road},c} + \Delta_{\text{term},c} + G_c, \qquad G_c \coloneqq u_c v_c^\top. \tag{A.63}$$

If a finer split is useful, R_c can itself be written as

$$R_c = \bar{J} + D_c + P_c^\mu + P_c^\lambda, \tag{A.64}$$

where

$$\bar{J} := \begin{bmatrix} -aI_N & bI_N \\ \gamma(I_N)_{1:(N-1),:} & -\varepsilon(I_N)_{1:(N-1),:} \\ \frac{\alpha}{\varphi}e_N^\top & -\frac{1+\beta(\sigma-1)}{(\sigma-1)\varphi}e_N^\top \end{bmatrix}, \tag{A.65}$$

$$D_c := \begin{bmatrix} S_x^c & 0 \\ 0 & (S_y^c)^{(N-1)} \\ 0 & 0 \end{bmatrix}, \tag{A.66}$$

$$P_c^\mu := \begin{bmatrix} aM_c & -bM_c \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \tag{A.67}$$

$$P_c^\lambda := \begin{bmatrix} 0 & 0 \\ -\gamma(\Lambda_c^\top)_{1:(N-1),:} & \varepsilon(\Lambda_c^\top)_{1:(N-1),:} \\ 0 & 0 \end{bmatrix}. \tag{A.68}$$

At the transport level, mode-specific congestion takes the AFW form

$$t_{ik,m}^c = \bar{t}_{ik,m}(\Xi_{ik,m}^c)^{\lambda_m}, \qquad t_{ik}^c = \Big(\sum_{m\in M(i,k)} (t_{ik,m}^c)^{-\nu}\Big)^{-1/\nu}, \qquad \pi_{ik}^{m,c} = \frac{(t_{ik,m}^c)^{-\nu}}{\sum_{n\in M(i,k)}(t_{ik,n}^c)^{-\nu}}. \tag{A.69}$$

Here Ξ^c collects the road-congestion states (mode-specific link flows) and T^c collects the terminal-congestion states relevant for transshipment and switching margins under model variant c. Hence

$$d\ln t_{ik}^c = \sum_{m\in M(i,k)} \pi_{ik}^{m,c}\, d\ln t_{ik,m}^c = \sum_{r\in\mathscr{S}_{\text{road}}} \chi_{ik;r}^{\text{road},c} d\ln \Xi_r^c + \sum_{t\in\mathscr{S}_{\text{term}}} \chi_{ik;t}^{\text{term},c} d\ln T_t^c, \tag{A.70}$$

where $\chi_{ik;r}^{\text{road},c}$ measures how road-congestion state r loads into the composite transport cost on link (i,k), and $\chi_{ik;t}^{\text{term},c}$ measures the analogous loading from terminal state t. Let

$$B_{t,c} := \frac{\partial G_c}{\partial \ln t^c}, \qquad Q_{\text{road},c} := \frac{\partial \ln \Xi^c}{\partial \zeta^\top}, \qquad Q_{\text{term},c} := \frac{\partial \ln T^c}{\partial \zeta^\top}, \tag{A.71}$$

In this notation, $B_{t,c}$ maps composite link-cost changes into the transformed equilibrium equations, while $Q_{\text{road},c}$ and $Q_{\text{term},c}$ map equilibrium responses into road- and terminal-congestion states. Define the loading operators $K_{\text{road},c}$ and $K_{\text{term},c}$ by

$$(K_{\text{road},c}\, u)_{ik} := \sum_{r\in\mathscr{S}_{\text{road}}} \chi_{ik;r}^{\text{road},c} u_r, \qquad (K_{\text{term},c}\, v)_{ik} := \sum_{t\in\mathscr{S}_{\text{term}}} \chi_{ik;t}^{\text{term},c} v_t. \tag{A.72}$$

Then the two additive congestion blocks in (A.61) are

$$\Delta_{\text{road},c} := B_{t,c}K_{\text{road},c}Q_{\text{road},c}, \qquad \Delta_{\text{term},c} := B_{t,c}K_{\text{term},c}Q_{\text{term},c}. \tag{A.73}$$

These are the exact Jacobian terms generated by road and terminal congestion feedback: $\Delta_{\text{road},c}$ records how equilibrium reacts when a mode-specific link improvement shifts traffic onto congested road states, while $\Delta_{\text{term},c}$ records the analogous feedback through terminals and switching margins.

Finally, because $L_i = W^{1/\rho} z_i$ and $\sum_i L_i = \bar{L}$,

$$d\ln W = -\frac{\rho}{\varphi}\sum_i \frac{L_i}{\bar{L}}\Big(d\ln x_i + \frac{\sigma}{\sigma-1} d\ln y_i\Big), \tag{A.74}$$

so the welfare row is

$$g^\top = \frac{\rho}{\varphi}\Big[\frac{L_1}{\bar{L}},\dots,\frac{L_N}{\bar{L}},\frac{\sigma}{\sigma-1}\frac{L_1}{\bar{L}},\dots,\frac{\sigma}{\sigma-1}\frac{L_N}{\bar{L}}\Big]. \tag{A.75}$$

A composite-link perturbation on link kl enters sparsely through

$$b_{kl} = B_t \mathbf{e}_{kl} = (1-\sigma) \begin{bmatrix} \mu_{kl} \mathbf{e}_k^x \\ \lambda_{kl} \mathbf{e}_l^y \end{bmatrix}, \tag{A.76}$$

with the second block omitted when required by the normalization row. Combining these ingredients with the implicit function theorem yields the grouped Jacobian decomposition (A.17), the finer reallocation-block split (A.64), and Propositions 3–6.

A.9 Derivations for Section 5: Implied Modal Wedge

We use the estimated elasticity $\hat{\eta}$ together with observed modal shipment flows to impute a reduced-form measure of generalized frictions that prevent adoption of non-road modes. We focus on the origin-side object used in the main text; the destination-side analogue is defined symmetrically.

Let road be the reference mode, indexed by 0, and let $m \in \{\text{rail}, \text{water-dom}, \text{water-for}\}$ denote an alternative mode. For each origin o, define total outgoing shipments by mode as

$$X_{o\cdot m} := \sum_{d} X_{odm}. \tag{A.77}$$

Under the gravity structure in the text, the observed road-relative modal flow ratio implies a revealed generalized-cost ratio

$$\rho_o(m\|0) := \left(\frac{X_{o\cdot m}}{X_{o\cdot 0}}\right)^{-1/\hat{\eta}}, \tag{A.78}$$

for all origins with $X_{o\cdot m} > 0$ and $X_{o\cdot 0} > 0$.

Next, let benchmark line-haul costs be linear in distance:

$$\bar{C}_m(D) = T_m D, \qquad \bar{C}_0(D) = T_0 D, \tag{A.79}$$

where T_m is the calibrated freight rate in dollars per TEU-mile for mode m. We define the implied multiplicative adoption wedge as

$$\Lambda_o(m\|0) := \rho_o(m\|0)\frac{T_0}{T_m}. \tag{A.80}$$

Equivalently,

$$\frac{\Lambda_o(m\|0)\bar{C}_m(D)}{\bar{C}_0(D)} = \rho_o(m\|0). \tag{A.81}$$

Thus, $\Lambda_o(m\|0) > 1$ means that mode m must face an additional generalized friction, above its calibrated engineering cost, for the model to reproduce the observed road-relative modal split.

Our preferred local summary is constructed on the common-support set

$$\mathcal{S}_m := \{o : X_{o\cdot m} > 0 \text{ and } X_{o\cdot 0} > 0\}, \tag{A.82}$$

using reference-mode flow weights

$$\omega_o := \frac{X_{o\cdot 0}}{\sum_{j\in\mathcal{S}_m} X_{j\cdot 0}}. \tag{A.83}$$

The average-local wedge is the weighted geometric mean of the origin-level wedges:

$$\ln \Lambda_m^{\text{avg}} = \sum_{o\in\mathcal{S}_m} \omega_o \ln \Lambda_o(m\|0), \qquad \Lambda_m^{\text{avg}} = \exp\big(\ln \Lambda_m^{\text{avg}}\big). \tag{A.84}$$

This is the common-support object reported in the main text.

We also report an exact national counterpart based on total modal flows:

$$\Lambda_m^{\text{nat}} = \left(\frac{\sum_o X_{o\cdot m}}{\sum_o X_{o\cdot 0}}\right)^{-1/\hat{\eta}} \frac{T_0}{T_m}. \tag{A.85}$$

The difference between Λ_m^{avg} and Λ_m^{nat} reflects both heterogeneity across origins and the role of common support.

Finally, for interpretability we convert the dimensionless wedge into a dollar-equivalent amount at benchmark distance D:

$$S_m(D) = \big(\Lambda_m - 1\big) T_m D. \tag{A.86}$$

This dollar-equivalent quantity should be interpreted as the money-equivalent generalized friction required to rationalize the observed modal split. It is not, in general, a literal terminal fee or one-for-one switching charge. Under linear per-mile benchmark costs, the multiplicative wedge Λ_m is invariant to the choice of D; only its dollar-equivalent representation $S_m(D)$ depends on the reporting distance.

A.10 Derivations for Section 5.1: Proofs of Proposition 1

This section presents the proof for Proposition 1. We proceed in two steps. In a first step we derive the change in the equilibrium conditions in terms of market access terms before then substitution the model specific elements.

A.10.1 Preliminaries

We can write equilibrium trade flows as,

$$X_{ij} = \tau_{ij}^{-\theta} \times \frac{\gamma_i}{\Pi_i^{-\theta}} \times \frac{\delta_j}{P_j^{-\theta}}$$

where γ_i and δ_j are cumulative flows out of and origin and into a destination, respectively, and Π_i and P_j are origin and destination market access terms. Given the recursive routing formulation, trade costs can be represented as:

$$\tau_{ij} = \left(\sum_{k \in \mathscr{F}(i)} \left(t_{ik} \tau_{kj} \right)^{-\theta} \right)^{-\frac{1}{\theta}}$$

And furthermore from the nested choice along the multi-layered graph, we have,

$$t_{ik}^{-\theta} = \left(\sum_{m \in \mathscr{M}(i,k)} \tilde{t}_{ik,m}^{-\eta} \right)^{\frac{\theta}{\eta}|}$$

In terms of market clearing and trade balance we have the following equilibrium conditions,

$$\gamma_i = \sum_j X_{ij}$$
$$\delta_i = \sum_j X_{ji}$$

A.10.2 Deriving the equilibrium equation

Starting with the first equilibrium condition we have:

$$\gamma_i = \sum_j X_{ij}$$

$$\gamma_i = \sum_j \tau_{ij}^{-\theta_j} \times \left(\frac{\gamma_i}{\Pi_i^{-\theta_j}} \right)^{\frac{\theta_j}{\theta_i}} \times \frac{\delta_j}{P_j^{-\theta_j}}$$

$$\Pi_i^{-\theta_i} = \sum_j \tau_{ij}^{-\theta_j} \times \Pi_i^{\theta_j \frac{\theta_j}{\theta_i} - \theta_i} (\gamma_i)^{\frac{\theta_j}{\theta_i} - 1} \times \frac{\delta_j}{P_j^{-\theta_j}}$$

$$\Pi_i^{-\theta_i} = \sum_j \tau_{ij}^{-\theta_j} \times \Pi_i^{\Delta_{i,j}} \times (\gamma_i)^{\frac{\theta_j}{\theta_i} - 1} \times \frac{\delta_j}{P_j^{-\theta_j}}$$

$$\Pi_i^{-\theta_i} = \tau_{ii}^{-\theta_i} \frac{\delta_i}{P_i^{-\theta_i}} + \sum_{j \neq i} \tau_{ij}^{-\theta_j} \Pi_i^{\Delta_{i,j}} (\gamma_i)^{\frac{\theta_j}{\theta_i} - 1} \frac{\delta_j}{P_j^{-\theta_j}}$$

$$\Pi_i^{-\theta_i} = \tau_{ii}^{-\theta_i} \frac{\delta_i}{P_i^{-\theta_i}} + \sum_{j \neq i} \left(\sum_{k \in \mathscr{B}(k)} t_{ik}^{-\theta_j} \tau_{kj}^{-\theta_j} \right) \Pi_i^{\Delta_{i,j}} (\gamma_i)^{\frac{\theta_j}{\theta_i} - 1} \frac{\delta_j}{P_j^{-\theta_j}}$$

$$\Pi_i^{-\theta_i} = \tau_{ii}^{-\theta_i} \frac{\delta_i}{P_i^{-\theta_i}} + \sum_{j \neq i} \left(\sum_{k \in \mathscr{F}(i)} t_{ik}^{-\theta_j} \tau_{kj}^{-\theta_j} \right) \Pi_i^{\Delta_{i,j}} (\gamma_i)^{\frac{\theta_j}{\theta_i} - 1} \frac{\delta_j}{P_j^{-\theta_j}}$$

$$\Pi_i^{-\theta_i} = \tau_{ii}^{-\theta_i} \frac{\delta_i}{P_i^{-\theta_i}} + \sum_{k \in \mathscr{F}(i)} t_{ik}^{-\theta_i} \left(-\tau_{ki}^{-\theta_i} \frac{\delta_i}{p_i^{-\theta_i}} + \sum_j t_{ik}^{\Delta\theta_j} \tau_{kj}^{-\theta_j} \Pi_i^{\Delta_{i,j}} (\gamma_i)^{\frac{\theta_j}{\theta_i} - 1} \frac{\delta_j}{P_j^{-\theta_j}} \right)$$

$$\Pi_i^{-\theta_i} = \left(t_{ii}^{-\theta_i} + \sum_{k \in \mathscr{F}(i)} (t_{ik} \tau_{ki})^{-\theta_i} \right) \frac{\delta_i}{P_i^{-\theta_i}} + \sum_{k \in \mathscr{F}(i)} t_{ik}^{-\theta_i} \left(-\tau_{ki}^{-\theta_i} \frac{\delta_i}{p_i^{-\theta_i}} + \sum_j t_{ik}^{\Delta\theta_j} \tau_{kj}^{-\theta_j} \Pi_i^{\Delta_{i,j}} (\gamma_i)^{\frac{\theta_j}{\theta_i} - 1} \frac{\delta_j}{P_j^{-\theta_j}} \right)$$

$$\Pi_i^{-\theta_i} = \left(t_{ii}^{-\theta_i} \right) \frac{\delta_i}{P_i^{-\theta_i}} + \sum_{k \in \mathscr{F}(i)} t_{ik}^{-\theta_i} \left(\sum_j t_{ik}^{\Delta\theta_j} \Pi_i^{\Delta_{i,j}} (\gamma_i)^{\frac{\theta_j}{\theta_i} - 1} \tau_{kj}^{-\theta_j} \frac{\delta_j}{P_j^{-\theta_j}} \right)$$

$$\Pi_i^{-\theta_i} = t_{ii}^{-\theta_i} \frac{\delta_i}{P_i^{-\theta_i}} + \sum_{k \in \mathscr{F}(i)} t_{ik}^{-\theta_i} \tilde{\Pi}_k^{-\theta_i}$$

where in the second line we defined, $\Pi_i^{\Delta_{i,j}} \equiv \Pi_i^{\theta_j \frac{\theta_j}{\theta_i} - \theta_i}$ and where $\tilde{\Pi}_k^{-\theta_i} = \left(\sum_j t_{ik}^{\Delta\theta_j} \Pi_i^{\Delta_{i,j}} (\gamma_i)^{\frac{\theta_j}{\theta_i} - 1} \tau_{kj}^{-\theta_j} \frac{\delta_j}{P_j^{-\theta_j}} \right)^{\frac{\theta_i}{\theta_j}}$ is the approximate price index that controls for differences in elasticities of substitution between neighboring nodes. And where in the fourth line we have used the fact that at the penultimate node the set of predecessor nodes of k ($\mathscr{B}(k)$) coincides with set of successor nodes of i ($\mathscr{F}(i)$).

Continuing with the second equilibrium condition,

$$\delta_j = \sum_i X_{ij}$$

$$\delta_j = \sum_i \tau_{ij}^{-\theta_j} \times \left(\frac{\gamma_i}{\Pi_i^{-\theta_j}} \right)^{\frac{\theta_j}{\theta_i}} \times \frac{\delta_j}{P_j^{-\theta_j}}$$

$$P_j^{-\theta_j} = \sum_i \tau_{ij}^{-\theta_j} \times \left(\frac{\gamma_i}{\Pi_i^{-\theta_j}} \right)^{\frac{\theta_j}{\theta_i}}$$

$$P_j^{-\theta_j} = \tau_{jj}^{-\theta_j} \frac{\gamma_j}{\Pi_j^{-\theta_j}} + \sum_{i \neq j} \tau_{ij}^{-\theta_j} \left(\frac{\gamma_i}{\Pi_i^{-\theta_j}} \right)^{\frac{\theta_j}{\theta_i}}$$

$$P_j^{-\theta_j} = \tau_{jj}^{-\theta_j} \frac{\gamma_j}{\Pi_j^{-\theta_j}} + \sum_{i \neq j} \left(\sum_{k \in \mathscr{B}(j)} \tau_{ik}^{-\theta_j} t_{kj}^{-\theta_j} \right) \left(\frac{\gamma_i}{\Pi_i^{-\theta_j}} \right)^{\frac{\theta_j}{\theta_i}}$$

$$P_j^{-\theta_j} = \tau_{jj}^{-\theta_j} \frac{\gamma_j}{\Pi_j^{-\theta}} + \sum_{k \in \mathscr{B}(j)} t_{kj}^{-\theta_j} \left(-\tau_{jk}^{-\theta_j} \frac{\gamma_j}{\Pi_j^{-\theta_j}} + \sum_i \tau_{ik}^{-\theta_j} \left(\frac{\gamma_i}{\Pi_i^{-\theta_j}} \right)^{\frac{\theta_j}{\theta_i}} \right)$$

$$P_j^{-\theta_j} = \left(t_{jj}^{-\theta_j} + \sum_{k \in \mathscr{B}(j)} \left(t_{jk} \tau_{kj} \right)^{-\theta_j} \right) \frac{\gamma_j}{\Pi_j^{-\theta_j}} + \sum_{k \in \mathscr{B}(j)} t_{kj}^{-\theta_j} \left(-\tau_{jk}^{-\theta_j} \frac{\gamma_j}{\Pi_j^{-\theta_j}} + \sum_i \tau_{ik}^{-\theta_j} \left(\frac{\gamma_i}{\Pi_i^{-\theta_j}} \right)^{\frac{\theta_j}{\theta_i}} \right)$$

$$P_j^{-\theta_j} = \left(t_{jj}^{-\theta_j} \right) \frac{\gamma_j}{\Pi_j^{-\theta_j}} + \sum_{k \in \mathscr{B}(j)} t_{kj}^{-\theta_j} \left(\sum_i \tau_{ik}^{-\theta_j} \left(\frac{\gamma_i}{\Pi_i^{-\theta_j}} \right)^{\frac{\theta_j}{\theta_i}} \right)$$

$$P_j^{-\theta_j} = t_{jj}^{-\theta_j} \frac{\gamma_j}{\Pi_j^{-\theta_j}} + \sum_{k \in \mathscr{B}(j)} t_{kj}^{-\theta_j} P_k^{-\theta_j}$$

where in the last line we used the definition of the recursive transport cost (Equation (3.1)), $\tau_{ii}^{-\theta} = \left(t_{ii}^{-\theta} + \sum_{k \in \mathscr{F}(i)} \left(t_{ik} \tau_{kj} \right)^{-\theta} \right)$. In what follows, we derive the counterfactual equation for the special case with symmetric substitution elasticities, ($\theta_i = \theta \quad \forall i$), i.e.

$$\Pi_i^{-\theta} = \left(t_{ii}^{-\theta} \right) \frac{\delta_i}{P_i^{-\theta}} + \sum_{k \in \mathscr{F}(i)} t_{ik}^{-\theta} \Pi_k^{-\theta}$$

$$P_j^{-\theta} = t_{jj}^{-\theta} \frac{\gamma_j}{\Pi_j^{-\theta}} + \sum_{k \in \mathscr{B}(j)} t_{kj}^{-\theta} P_k^{-\theta}$$

We can make the relationship between the mode-specific transportation cost and the spatial aggregate equilibrium more explicit by substitution the definition of the edge-specific transport cost in terms of the multi-layered mode specific transport cost, i.e.

$$P_i^{-\theta} = t_{ii}^{-\theta} \frac{\gamma_i}{\Pi_i^{-\theta}} + \sum_{k \in \mathscr{F}(j)} \left(\sum_{m \in \mathscr{M}(k,i)} \tilde{t}_{ki,m}^{-\eta} \right)^{\frac{\theta}{\eta}|} P_k^{-\theta}$$

$$= t_{ii}^{-\theta} \frac{\gamma_i}{\Pi_i^{-\theta}} + \sum_{k \in \mathscr{F}(j)} \left(\sum_{m \in \mathscr{M}(k,i)} \tilde{t}_{ki,m}^{-\eta} P_k^{-\eta} \right)^{\frac{\theta}{\eta}|}$$

$$\Pi_i^{-\theta} = t_{ii}^{-\theta}\frac{\delta_i}{P_i^{-\theta}} + \sum_{k\in\mathscr{F}(i)}\left(\sum_{m\in\mathscr{M}(i,k)}\tilde{t}_{ik,m}^{-\eta}\right)^{\frac{\theta}{\eta}|}\Pi_k^{-\theta}$$
$$= t_{ii}^{-\theta}\frac{\delta_i}{P_i^{-\theta}} + \sum_{k\in\mathscr{F}(i)}\left(\sum_{m\in\mathscr{M}(i,k)}\tilde{t}_{ik,m}^{-\eta}\Pi_k^{-\eta}\right)^{\frac{\theta}{\eta}|}$$

where we notice that the equilibrium equations then inherit a the structure of nested choice models directly and where the modal substitution parameter than governs the impact of mode-specific changes on aggregate endogenous variables.

A.10.3 Deriving the equilibrium system in changes

Start with (recursive) equilibrium condition from the previous subsection,

$$\Pi_i^{-\theta} = t_{ii}^{-\theta}\frac{\delta_i}{P_i^{-\theta}} + \sum_{k\in\mathscr{F}(i)} t_{ik}^{-\theta}\Pi_k^{-\theta}$$

$$P_i^{-\theta} = t_{ii}^{-\theta}\frac{\gamma_i}{\Pi_i^{-\theta}} + \sum_{k\in\mathscr{F}(j)} t_{ki}^{-\theta}P_k^{-\theta}$$

Expressed in changes,

$$\hat{\Pi}_i^{-\theta} = \left(\frac{t_{ii}^{-\theta}\frac{\delta_i}{P_i^{-\theta}}}{t_{ii}^{-\theta}\frac{\delta_i}{P_i^{-\theta}} + \sum_{k\in\mathscr{F}(i)} t_{ik}^{-\theta}\Pi_k^{-\theta}}\right)\frac{\hat{\delta}_i}{\hat{P}_i^{-\theta}} + \sum_{k\in\mathscr{F}(i)}\left(\frac{t_{ik}^{-\theta}\Pi_k^{-\theta}}{t_{ii}^{-\theta}\frac{\delta_i}{P_i^{-\theta}} + \sum_{k\in\mathscr{F}(i)} t_{ik}^{-\theta}\Pi_k^{-\theta}}\right)\hat{t}_{ik}^{-\theta}\hat{\Pi}_k^{-\theta}$$

and,

$$\hat{P}_i^{-\theta} = \left(\frac{t_{ii}^{-\theta}\frac{\gamma_i}{\Pi_i^{-\theta}}}{t_{ii}^{-\theta}\frac{\gamma_i}{\Pi_i^{-\theta}} + \sum_{k\in\mathscr{F}(j)} t_{ki}^{-\theta}P_k^{-\theta}}\right)\frac{\hat{\gamma}_i}{\hat{\Pi}_i^{-\theta}} + \sum_{k\in\mathscr{F}(i)}\left(\frac{t_{ki}^{-\theta}P_k^{-\theta}}{t_{ii}^{-\theta}\frac{\gamma_i}{\Pi_i^{-\theta}} + \sum_{k\in\mathscr{F}(i)} t_{ki}^{-\theta}P_k^{-\theta}}\right)\hat{t}_{ki}^{-\theta}\hat{P}_k^{-\theta}$$

Multiplying both numerator and denominator by their appropriate market access terms, we obtain,

$$\hat{\Pi}_i^{-\theta} = \left(\frac{\delta_i}{\delta_i + \sum_{k\in\mathscr{F}(i)} t_{ik}^{-\theta}P_k^{-\theta}\Pi_j^{-\theta}}\right)\frac{\hat{\delta}_i}{\hat{P}_i^{-\theta}} + \sum_{k\in\mathscr{F}(i)}\left(\frac{t_{ik}^{-\theta}\Pi_j^{-\theta}P_i^{-\theta}}{\delta_i + \sum_{k\in\mathscr{F}(i)} t_{ik}^{-\theta}P_k^{-\theta}\Pi_j^{-\theta}}\right)\hat{t}_{ik}^{-\theta}\hat{\Pi}_k^{-\theta}$$

$$\hat{P}_i^{-\theta} = \left(\frac{\gamma_i}{\gamma_i + \sum_{k\in\mathscr{F}(i)} t_{ki}^{-\theta}P_k^{-\theta}\Pi_i^{-\theta}}\right)\frac{\hat{\gamma}_i}{\hat{\Pi}_i^{-\theta}} + \sum_{k\in\mathscr{F}(i)}\left(\frac{t_{ki}^{-\theta}P_k^{-\theta}\Pi_i^{-\theta}}{\gamma_i + \sum_{k\in\mathscr{F}(i)} t_{ki}^{-\theta}P_k^{-\theta}\Pi_i^{-\theta}}\right)\hat{t}_{ki}^{-\theta}\hat{P}_k^{-\theta}$$

Simplifying we obtain,

$$\hat{\Pi}_i^{-\theta} = \left(\frac{\delta_i}{\delta_i + \sum_{k \in \mathscr{F}(i)} \Xi_{ik}} \right) \frac{\hat{\delta}_i}{\hat{P}_i^{-\theta}} + \sum_{k \in \mathscr{F}(i)} \left(\frac{\Xi_{ik}}{\delta_i + \sum_{k \in \mathscr{F}(i)} \Xi_{ik}} \right) \hat{t}_{ik}^{-\theta} \hat{\Pi}_k^{-\theta}$$

$$\hat{P}_i^{-\theta} = \left(\frac{\gamma_i}{\gamma_i + \sum_{k \in \mathscr{F}(i)} \Xi_{ki}} \right) \frac{\hat{\gamma}_i}{\hat{\Pi}_i^{-\theta}} + \sum_{k \in \mathscr{F}(i)} \left(\frac{\Xi_{ki}}{\gamma_i + \sum_{k \in \mathscr{F}(i)} \Xi_{ki}} \right) \hat{t}_{ki}^{-\theta} \hat{P}_k^{-\theta}$$

Next, we proceed by deriving the nested transport equilibrium at the edge-level. Recall, that given the nested choice above the changes in aggregate transport cost is given by,

$$\hat{t}_{ik}^{-\theta} = \left(\sum_{m \in \mathscr{M}(i,k)} \frac{\left(t'_{ik,m} \right)^{-\eta}}{t_{ik}^{-\eta}} \right)^{\frac{\theta}{\eta}}$$

Multiplying with the appropriate market access terms, we obtain,

$$\hat{t}_{ik}^{-\theta} = \left(\sum_{m \in \mathscr{M}(i,k)} \frac{\Xi_{ik.m}}{\Xi_{ik}} \hat{t}_{ik,m}^{-\eta} \right)^{\frac{\theta}{\eta}}$$

For the changes in transport cost we have,

$$\hat{t}_{kl,1} = \hat{\hat{t}}_{kl,0}^{\frac{1}{1+\eta\lambda_0}} \times \hat{t}_{kl}^{\frac{\lambda_0(\eta-\theta)}{1+\eta\lambda_0}} \times \hat{P}_k^{\frac{-\theta\lambda_0}{1+\eta\lambda_0}} \times \hat{\Pi}_l^{\frac{-\theta\lambda_0}{1+\eta\lambda_0}}$$

And second, we have for any non-primary transport mode,

$$\hat{t}_{kl,m} = \hat{\hat{s}}_{kk,m} \hat{\tau}_{kl,m} \hat{\hat{s}}_{ll,m} \hat{P}_k^{-\theta\lambda_m} \hat{\Pi}_{k,m}^{-\theta\lambda_m} \hat{\Pi}_l^{-\theta\lambda_m} \hat{P}_{l,m}^{-\theta\lambda_m}$$

For the trade model here we have,

$$\delta_i = E_i$$
$$\gamma_i = Y_i$$

And for the price indices in changes we have,

$$\hat{P}_i = \hat{y}_i \hat{l}_i^{\beta-1} \hat{W}^{-1}$$
$$\hat{\Pi}_i = \hat{l}_i^{\alpha+1} \hat{y}_i^{-\frac{\theta+1}{\theta}}$$

and furthermore, we have the mode specific market access terms,

$$\hat{P}_{l,m}^{-\theta} \equiv \sum_{k \in \mathscr{B}(l)} \frac{\Xi_{kl,m}}{\sum_{k \in \mathscr{B}(l)} \Xi_{kl,m}} \hat{t}_{kl,m}^{-\eta} \times \hat{t}_{kl}^{\eta-\theta} \times \hat{P}_k^{-\theta}$$

$$\hat{\Pi}_{k,m}^{-\theta} \equiv \sum_{l \in \mathscr{F}(k)} \frac{\Xi_{kl,m}}{\sum_{l \in \mathscr{F}(k)} \Xi_{kl,m}} \times \hat{t}_{kl,m}^{-\eta} \times \hat{t}_{kl}^{\eta-\theta} \times \hat{\Pi}_l^{-\theta}$$

Given an initial guess for the market access terms, we can solve for the aggregate transport costs, $\hat{t}_{ik}^{-\theta}\left(\hat{\Pi}, \hat{P}, \hat{t}_{ik,1}, \hat{t}_{ik,m}\right)$, by iteratively updating the following equation until convergence,

$$\begin{aligned}
\hat{t}_{ik}^{-\theta} &= \left(\frac{\Xi_{ik.0}}{\Xi_{ik}} \hat{t}_{ik,0}^{-\eta} + \sum_{m \neq 0} \frac{\Xi_{ik.m}}{\Xi_{ik}} \hat{t}_{ik,m}^{\eta} \right)^{\frac{\theta}{\eta}} \\
&= \left(\frac{\Xi_{ik.0}}{\Xi_{ik}} \left[\hat{\hat{t}}_{ik,0}^{\frac{1}{+\eta\eta\lambda_0}} \times \hat{t}_{ik}^{\frac{\lambda_0(\eta-\theta)}{1+\eta\lambda_0}} \times \hat{P}_i^{\frac{-\theta\lambda_0}{1+\eta\lambda_0}} \times \hat{\Pi}_k^{\frac{-\theta\lambda_0}{1+\eta\lambda_0}} \right]^{-\eta} + \sum_{m \neq 0} \frac{\Xi_{ik.m}}{\Xi_{ik}} \left[\hat{\hat{s}}_{ii,m} \hat{\tau}_{ik,m} \hat{\hat{s}}_{kk,m} \hat{P}_i^{-\theta\lambda_m} \hat{\Pi}_{i,m}^{-\theta\lambda_m} \hat{\Pi}_k^{-\theta\lambda_m} \hat{P}_{k,m}^{-\theta\lambda_m} \right]^{-\eta} \right)^{\frac{\theta}{\eta}}
\end{aligned}$$

Writing out and substituting for the transport cost terms,

$$\hat{P}_i^{-\theta}\hat{\Pi}_i^{-\theta} = \left(\frac{\delta_i}{\delta_i + \sum_{k\in\mathscr{F}(i)}\Xi_{ik}}\right)\hat{\delta}_i + \sum_{k\in\mathscr{F}(i)}\left(\frac{\Xi_{ik}}{\delta_i + \sum_{k\in\mathscr{F}(i)}\Xi_{ik}}\right)\hat{t}_{ik}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{\tilde{t}}_{ik,0},\hat{\tilde{t}}_{ik,m}\right)\hat{P}_i^{-\theta}\hat{\Pi}_k^{-\theta}$$

$$\hat{P}_i^{-\theta}\hat{\Pi}_i^{-\theta} = \left(\frac{\gamma_i}{\gamma_i + \sum_{k\in\mathscr{F}(i)}\Xi_{ki}}\right)\hat{\gamma}_i + \sum_{k\in\mathscr{F}(i)}\left(\frac{\Xi_{ki}}{\gamma_i + \sum_{k\in\mathscr{F}(i)}\Xi_{ki}}\right)\hat{t}_{ki}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{\tilde{t}}_{ki,0},\hat{\tilde{t}}_{ki,m}\right)\hat{P}_k^{-\theta}\hat{\Pi}_i^{-\theta}$$

Substituting the price indices in changes, we get,

$$\begin{aligned}\left(\hat{l}_i^{\alpha+1}\hat{y}_i^{-\frac{\theta+1}{\theta}}\right)^{-\theta}\left(\hat{y}_i\hat{l}_i^{\beta-1}\hat{W}^{-1}\right)^{-\theta} &= \left(\frac{E_i}{E_i + \sum_{k\in\mathscr{F}(i)}\Xi_{ik}}\right)\hat{y}_i \\ &\quad + \sum_{k\in\mathscr{F}(i)}\left(\frac{\Xi_{ik}}{E_i + \sum_{k\in\mathscr{F}(i)}\Xi_{ik}}\right)\hat{t}_{ik}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{ik,0},\hat{t}_{ik,m}\right)\left(\hat{y}_i\hat{l}_i^{\beta-1}\hat{W}^{-1}\right)^{-\theta}\left(\hat{l}_k^{\alpha+1}\hat{y}_k^{-\frac{\theta+1}{\theta}}\right)^{-\theta} \\ \left(\hat{l}_i^{\alpha+1}\hat{y}_i^{-\frac{\theta+1}{\theta}}\right)^{-\theta}\left(\hat{y}_i\hat{l}_i^{\beta-1}\hat{W}^{-1}\right)^{-\theta} &= \left(\frac{Y_i}{Y_i + \sum_{k\in\mathscr{F}(i)}\Xi_{ki}}\right)\hat{y}_i \\ &\quad + \sum_{k\in\mathscr{F}(i)}\left(\frac{\Xi_{ki}}{Y_i + \sum_{k\in\mathscr{F}(i)}\Xi_{ki}}\right)\hat{t}_{ki}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{ki,0},\hat{t}_{ki,m}\right)\left(\hat{y}_k\hat{l}_k^{\beta-1}\hat{W}^{-1}\right)^{-\theta}\left(\hat{l}_i^{\alpha+1}\hat{y}_i^{-\frac{\theta+1}{\theta}}\right)^{-\theta}\end{aligned}$$

Simplifying and defining, $\hat{\chi} = \hat{W}^{-\theta}$, we have,

$$\begin{aligned}\hat{y}_i^{1+\theta}\hat{l}_i^{-\theta(1+\alpha)} &= \hat{\chi}\left(\frac{E_i}{E_i + \sum_{k\in\mathscr{F}(i)}\Xi_{ik}}\right)\hat{y}_i^{1+\theta}\hat{l}_i^{-\theta(1-\beta)} + \sum_{k\in\mathscr{F}(i)}\left(\frac{\Xi_{ik}}{E_i + \sum_{k\in\mathscr{F}(i)}\Xi_{ik}}\right)\hat{t}_{ik}^{-\theta}\hat{y}_k^{1+\theta}\hat{l}_k^{-\theta(1+\alpha)} \\ \hat{y}_i^{-\theta}\hat{l}_i^{\theta(1-\beta)} &= \hat{\chi}\left(\frac{Y_i}{Y_i + \sum_{k\in\mathscr{F}(i)}\Xi_{ki}}\right)\hat{y}_i^{-\theta}\hat{l}_i^{\theta(1+\alpha)} + \sum_{k\in\mathscr{F}(i)}\left(\frac{\Xi_{ki}}{Y_i + \sum_{k\in\mathscr{F}(i)}\Xi_{ki}}\right)\hat{t}_{ki}^{-\theta}\hat{y}_k^{-\theta}\hat{l}_k^{\theta(1-\beta)}\end{aligned}$$

A.11 Derivations for Section 5.3: Predicted vs Observed Transport Mode-level Trade Flows

For aggregate traffic we have,

$$\mathbf{\Xi} = \mathbf{PA}\Pi$$

Inverting,

$$\begin{aligned}\mathbf{\Xi} &= \mathbf{PA}\Pi \iff \\ \mathbf{A} &= \mathbf{P}^{-1}\mathbf{\Xi}\Pi^{-1}\end{aligned}$$

For the traffic equations we have,

$$\mathbf{\Xi}_m = \mathbf{PT_m}\Pi$$

Inverting,

$$\begin{aligned}\mathbf{\Xi}_m &= \mathbf{PT_m}\Pi \iff \\ \mathbf{T}_m &= \mathbf{P}^{-1}\mathbf{\Xi}_m\Pi^{-1}\end{aligned}$$

Bilateral mode specific shares are given by,

$$\begin{aligned}\pi_{ij,k,m} &= \frac{(w_i/A_i)^{-\theta}(t_{ik}\tau_{ki})^{-\theta}}{\sum_{i\in\mathscr{N}}(w_i/A_i)^{-\theta}\sum_{k'\in\mathscr{F}(i)}(t_{ik'}\tau_{k'i})^{-\theta}}\frac{\tilde{t}_{ik,m}^{-\eta}}{t_{ik}} \\ &= \pi_{ij,k}\times\pi_{ij,k}^m\end{aligned}$$

$$X_{ij} = \left(\tau_{ij}\right)^{-\theta}\times\frac{\gamma_i}{\Pi_i^{-\theta}}\times\frac{\delta_j}{P_j^{-\theta}}$$

$$X_{ij} = \left(\sum_{k\in\mathcal{N}} t_{ik}^{-\theta}\tau_{kj}^{-\theta}\right) \times \frac{\gamma_i}{\Pi_i^{-\theta}} \times \frac{\delta_j}{P_j^{-\theta}}$$
$$= \sum_{k\in\mathcal{N}} \underbrace{t_{ik}^{-\theta} \times \tau_{kj}^{-\theta} \times \frac{\gamma_i}{\Pi_i^{-\theta}} \times \frac{\delta_j}{P_j^{-\theta}}}_{\equiv X_{ij,k}}$$

$$X_{ij} = \left(\sum_{k\in\mathcal{N}}\sum_{l\in\mathcal{N}(k)} \tau_{ik}^{-\theta} t_{kl}^{-\theta}\tau_{lj}^{-\theta}\right) \times \frac{\gamma_i}{\Pi_i^{-\theta}} \times \frac{\delta_j}{P_j^{-\theta}}$$
$$= \sum_{k\in\mathcal{N}}\sum_{l\in\mathcal{N}(k)} \underbrace{\tau_{ik}^{-\theta} t_{kl}^{-\theta}\tau_{lj}^{-\theta} \times \frac{\gamma_i}{\Pi_i^{-\theta}} \times \frac{\delta_j}{P_j^{-\theta}}}_{\equiv X_{ij,kl}}$$

The share of flows between ij, along link kl and along mode m,

$$X_{ij,kl} = \left(\tau_{ik} t_{kl} \tau_{lj}\right)^{-\theta} \times \frac{\gamma_i}{\Pi_i^{-\theta}} \times \frac{\delta_j}{P_j^{-\theta}}$$

$$X_{ij,kl,m} = \frac{\tilde{t}_{kl,m}^{-\eta}}{t_{kl}^{-\eta}} \times \left(\tau_{ik} t_{kl} \tau_{lj}\right)^{-\theta} \times \frac{\gamma_i}{\Pi_i^{-\theta}} \times \frac{\delta_j}{P_j^{-\theta}}$$

sum over kl to obtain

$$X_{ij,m} = \sum_{k\in\mathbf{N}}\sum_{l\in\mathcal{N}(k)} \frac{\tilde{t}_{kl,m}^{-\eta}}{t_{kl}^{-\eta}} \times \left(\tau_{ik} t_{kl} \tau_{lj}\right)^{-\theta} \times \frac{\gamma_i}{\Pi_i^{-\theta}} \times \frac{\delta_j}{P_j^{-\theta}}$$
$$X_{ij,m} = \frac{\gamma_i}{\Pi_i^{-\theta}} \frac{\delta_j}{P_j^{-\theta}} \sum_{k\in\mathbf{N}} \tau_{ik}^{-\theta} \sum_{l\in\mathcal{N}(k)} \tau_{lj}^{-\theta} \times \frac{\tilde{t}_{kl,m}^{-\eta}}{t_{kl}^{-\eta}} t_{kl}^{-\theta}$$
$$X_{ij,m} = \frac{\gamma_i}{\Pi_i^{-\theta}} \frac{\delta_j}{P_j^{-\theta}} \sum_{k\in\mathbf{N}} \tau_{ik}^{-\theta} \sum_{l\in\mathcal{N}(k)} \tau_{lj}^{-\theta} \times \frac{\tilde{t}_{kl,m}^{-\eta}}{t_{kl}^{\theta-\eta}}$$

In matrix notation, we have,

$$\mathbf{X}_m = \left(\frac{Y}{\Pi}\right)(\mathbf{I}-\mathbf{A})^{-1}\mathbf{T}_m(\mathbf{I}-\mathbf{A})^{-1}\left(\frac{E}{P}\right)$$

Derivations,

$$
\begin{aligned}
\mathbf{X}_m &= \left(\frac{Y}{\Pi}\right)(\mathbf{I}-\mathbf{A})^{-1}\mathbf{T}_m(\mathbf{I}-\mathbf{A})^{-1}\left(\frac{E}{\mathbf{P}}\right) \Longleftrightarrow \\
\mathbf{X}_m &= \left(\frac{Y}{\Pi}\right)(\mathbf{I}-\mathbf{A})^{-1}\mathbf{P}^{-1}\mathbf{\Xi}_m\Pi^{-1}(\mathbf{I}-\mathbf{A})^{-1}\left(\frac{E}{\mathbf{P}}\right) \Longleftrightarrow \\
\mathbf{X}_m &= \left(\frac{Y}{\Pi}\right)\left(\mathbf{I}-\mathbf{P}^{-1}\Xi\Pi^{-1}\right)^{-1}\mathbf{P}^{-1}\mathbf{\Xi}_m\Pi^{-1}\left(\mathbf{I}-\mathbf{P}^{-1}\Xi\Pi^{-1}\right)^{-1}\left(\frac{E}{\mathbf{P}}\right) \Longleftrightarrow \\
\mathbf{X}_m^{-1} &= \left(\frac{E}{\mathbf{P}}\right)^{-1}\left(\mathbf{I}-\mathbf{P}^{-1}\Xi\Pi^{-1}\right)\Pi\mathbf{\Xi}_m^{-1}\mathbf{P}\left(\mathbf{I}-\mathbf{P}^{-1}\Xi\Pi^{-1}\right)\left(\frac{Y}{\Pi}\right)^{-1} \Longleftrightarrow \\
\mathbf{X}_m^{-1} &= \left(\left(\frac{E}{\mathbf{P}}\right)^{-1}-\left(\frac{E}{\mathbf{P}}\right)^{-1}\mathbf{P}^{-1}\Xi\Pi^{-1}\right)\Pi\mathbf{\Xi}_m^{-1}\mathbf{P}\left(\left(\frac{Y}{\Pi}\right)^{-1}-\mathbf{P}^{-1}\Xi\Pi^{-1}\left(\frac{Y}{\Pi}\right)^{-1}\right) \Longleftrightarrow \\
\mathbf{X}_m^{-1} &= \left(\left(\frac{E}{\mathbf{P}}\right)^{-1}\Pi-\left(\frac{E}{\mathbf{P}}\right)^{-1}\mathbf{P}^{-1}\Xi\Pi^{-1}\Pi\right)\mathbf{\Xi}_m^{-1}\left(\mathbf{P}\left(\frac{Y}{\Pi}\right)^{-1}-\mathbf{P}\mathbf{P}^{-1}\Xi\Pi^{-1}\left(\frac{Y}{\Pi}\right)^{-1}\right) \Longleftrightarrow \\
\mathbf{X}_m^{-1} &= E^{-1}(\mathbf{P}\Pi-\Xi)\mathbf{\Xi}_m^{-1}(\mathbf{P}\Pi-\Xi)Y^{-1} \Longleftrightarrow \\
\mathbf{X}_m &= Y(\mathbf{P}\Pi-\Xi)^{-1}\mathbf{\Xi}_m(\mathbf{P}\Pi-\Xi)^{-1}E
\end{aligned}
$$

Finally, we need an expression for $\mathbf{P}\Pi$,

$$\Pi P = Y + \Xi^T 1$$

and,

$$\Pi P = E + \Xi 1$$

We have,

$$\mathbf{X}_m = (Y)\left(\text{diag}\left(\frac{1}{2}(E+Y)+\frac{1}{2}\left(\Xi 1+\Xi^T 1\right)\right)-\Xi\right)^{-1}\mathbf{\Xi}_m\left(\text{diag}\left(\frac{1}{2}(E+Y)+\frac{1}{2}\left(\Xi 1+\Xi^T 1\right)\right)-\Xi\right)^{-1}(E)$$

We finally obtain,

$$X_{ij,m} = Y_i \times E_j \times \sum_{(k,l)}\left[\mathbf{D}^X-\Xi\right]^{-1}_{ik}\left[\mathbf{\Xi}_m\right]_{kl}\left[\mathbf{D}^X-\Xi\right]^{-1}_{lj}, \tag{A.87}$$

where $\mathbf{D}^X$ is the diagonal matrix with diagonal elements given by $\left(\frac{1}{2}(E+Y)+\frac{1}{2}\left(\Xi 1+\Xi^T 1\right)\right)_{ii}$

A.12 Derivations for Section 5.1: International Trade

In this subsection, we extend the model to allow for both domestic locations $i, j \in \mathscr{D}$ and foreign locations $l \in \mathscr{F}$. Domestic locations consumer from both domestic and foreign locations, while foreign locations are modeled to only trade with domestic locations. Furthermore, we make a small open economy assumption such that changes in the domestic economy do not affect the labor allocations in the foreign economy. For the domestic economy we have the following conditions,

$$\gamma_i = \sum_{j \in \mathscr{D}} X_{ij} + \sum_{l \in \mathscr{F}} X_{il}$$
$$\delta_i = \sum_{j \in \mathscr{D}} X_{ji} + \sum_{l \in \mathscr{F}} X_{li}$$

Welfare equalization domestically implies,

$$P_i = \frac{w_i u_i}{W} \iff$$
$$P_i = Y_i \bar{u}_i L_i^{\beta-1} W^{-1} \implies$$
$$\hat{P}_i = \hat{y}_i \hat{l}_i^{\beta-1} \hat{W}^{-1}$$

and

$$\Pi_i = A_i L_i Y_i^{-\frac{\theta+1}{\theta}} \iff$$
$$\Pi_i = A_i L_i^{\alpha+1} Y_i^{-\frac{\partial+1}{\theta}} \implies$$
$$\hat{\Pi}_i = \hat{l}_i^{\alpha+1} \hat{y}_i^{-\frac{\theta+1}{\theta}}$$

For the foreign economy we have that market clearing implies,

$$\gamma_l = \sum_j X_{lj} \quad \forall l \in \mathscr{F}$$
$$\delta_l = \sum_j X_{jl} \quad \forall l \in \mathscr{F}$$

For the price indices for foreign locations we obtain,

$$P_l = \frac{w_l u_l}{W_l} \iff$$
$$P_l = Y_l \bar{u}_l L_l^{\beta-1} W_l^{-1} \implies$$
$$\hat{P}_l = \hat{y}_l \hat{W}_l^{-1}$$

and

$$\Pi_l = A_l L_l Y_l^{-\frac{\theta+1}{\theta}} \iff$$
$$\Pi_l = A_l L_l^{\alpha+1} Y_l^{-\frac{\theta+1}{\theta}} \implies$$
$$\hat{\Pi}_l = \hat{y}_l^{-\frac{\theta+1}{\theta}}$$

Considering now that the economy is situated on a graph with both foreign and domestic nodes, i.e. $\mathscr{G} \equiv (\mathscr{N}, \mathscr{L})$, where $\mathscr{N}$ is the union of domestic, $\mathscr{N}_D$, and foreign nodes, $\mathscr{N}_F$, and equivalently we define domestic and foreign successor nodes, $k \in \mathscr{F}_D(i)$, and $l \in \mathscr{F}_F(i)$.

Re-writing the equilibrium system for both foreign and domestic locations, defining domestic and foreign neighboring

notes

$$\hat{P}_i^{-\theta}\hat{\Pi}_i^{-\theta} = \left(\frac{E_i}{E_i + \sum_{k\in\mathscr{F}(i)}\Xi_{ik}}\right)\hat{\delta}_i + \sum_{k\in\mathscr{F}_D(i)}\left(\frac{\Xi_{ik}}{E_i + \sum_{k\in\mathscr{F}(i)}\Xi_{ik}}\right)\hat{t}_{ik}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{ik,1},\hat{t}_{ik,m}\right)\hat{\Pi}_k^{-\theta}\hat{P}_i^{-\theta} + \sum_{l\in\mathscr{F}_F(i)}\left(\frac{\Xi_{il}}{E_i + \sum_{k\in\mathscr{F}(i)}\Xi_{ik}}\right)\hat{t}_{il}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{ik,1},\hat{t}_{ik,m}\right)\hat{P}_i^{-\theta}\hat{\Pi}_l^{-\theta}$$

$$\hat{P}_i^{-\theta}\hat{\Pi}_i^{-\theta} = \left(\frac{Y_i}{Y_i + \sum_{k\in\mathscr{F}(i)}\Xi_{ki}}\right)\hat{Y}_i + \sum_{k\in\mathscr{F}_D(i)}\left(\frac{\Xi_{ki}}{Y_i + \sum_{k\in\mathscr{F}(i)}\Xi_{ki}}\right)\hat{t}_{ki}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{ki,1},\hat{t}_{ki,m}\right)\hat{P}_k^{-\theta}\hat{\Pi}_i^{-\theta} + \sum_{l\in\mathscr{F}_F(i)}\left(\frac{\Xi_{li}}{Y_i + \sum_{k\in\mathscr{F}(i)}\Xi_{ki}}\right)\hat{t}_{li}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{ki,1},\hat{t}_{ki,m}\right)\hat{P}_l^{-\theta}\hat{\Pi}_i^{-\theta}$$

And similarly for foreign locations, we have,

$$\hat{\Pi}_l^{-\theta}\hat{P}_l^{-\theta} = \left(\frac{E_l}{E_l + \sum_{k\in\mathscr{F}_D(l)}\Xi_{lk}}\right)\hat{y}_l + \sum_{k\in\mathscr{F}_D(l)}\left(\frac{\Xi_{lk}}{E_l + \sum_{k\in\mathscr{F}_D(l)}\Xi_{lk}}\right)\hat{t}_{lk}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{lk,1},\hat{t}_{lk,m}\right)\hat{P}_l^{-\theta}\hat{\Pi}_k^{-\theta}$$

$$\hat{P}_l^{-\theta}\hat{\Pi}_l^{-\theta} = \left(\frac{Y_l}{Y_l + \sum_{k\in\mathscr{F}_D(l)}\Xi_{kl}}\right)\hat{y}_l + \sum_{k\in\mathscr{F}_D(l)}\left(\frac{\Xi_{kl}}{Y_l + \sum_{k\in\mathscr{F}_D(l)}\Xi_{kl}}\right)\hat{t}_{kl}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{kl,1},\hat{t}_{kl,m}\right)\hat{P}_k^{-\theta}\hat{\Pi}_l^{-\theta}$$

Substituting,

$$\left(\hat{l}_i^{\alpha+1}\hat{y}_i^{-\frac{\theta+1}{\theta}}\right)^{-\theta}\left(\hat{y}_i\hat{l}_i^{\beta-1}\hat{W}^{-1}\right)^{-\theta} = \left(\frac{E_i}{E_i + \sum_{k\in\mathscr{F}(i)}\Xi_{ik}}\right)\hat{y}_i + \sum_{k\in\mathscr{F}_D(i)}\left(\frac{\Xi_{ik}}{E_i + \sum_{k\in\mathscr{F}(i)}\Xi_{ik}}\right)\hat{t}_{ik}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{ik,1},\hat{t}_{ik,m}\right)\left(\hat{y}_i\hat{l}_i^{\beta-1}\hat{W}^{-1}\right)^{-\theta}\left(\hat{l}_k^{\alpha+1}\hat{y}_k^{-\frac{\theta+1}{\theta}}\right)^{-\theta} + \sum_{l\in\mathscr{F}_F(i)}\left(\frac{\Xi_{il}}{E_i + \sum_{k\in\mathscr{F}(i)}\Xi_{ik}}\right)\hat{t}_{il}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{ik,1},\hat{t}_{ik,m}\right)\left(\hat{y}_i\hat{l}_i^{\beta-1}\hat{W}^{-1}\right)^{-\theta}\left(\hat{y}_l^{-\frac{\theta+1}{\theta}}\right)^{-\theta}$$

$$\left(\hat{l}_i^{\alpha+1}\hat{y}_i^{-\frac{\theta+1}{\theta}}\right)^{-\theta}\left(\hat{y}_i\hat{l}_i^{\beta-1}\hat{W}^{-1}\right)^{-\theta} = \left(\frac{Y_i}{Y_i + \sum_{k\in\mathscr{F}(i)}\Xi_{ki}}\right)\hat{y}_i + \sum_{k\in\mathscr{F}_D(i)}\left(\frac{\Xi_{ki}}{Y_i + \sum_{k\in\mathscr{F}(i)}\Xi_{ki}}\right)\hat{t}_{ki}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{ki,1},\hat{t}_{ki,m}\right)\left(\hat{y}_k\hat{l}_k^{\beta-1}\hat{W}^{-1}\right)^{-\theta}\left(\hat{l}_i^{\alpha+1}\hat{y}_i^{-\frac{\theta+1}{\theta}}\right)^{-\theta} + \sum_{l\in\mathscr{F}_F(i)}\left(\frac{\Xi_{li}}{Y_i + \sum_{k\in\mathscr{F}(i)}\Xi_{ki}}\right)\hat{t}_{li}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{ki,1},\hat{t}_{ki,m}\right)\left(\hat{y}_l\hat{W}_l^{-1}\right)^{-\theta}\left(\hat{l}_i^{\alpha+1}\hat{y}_i^{-\frac{\theta+1}{\theta}}\right)^{-\theta}$$

$$\left(\hat{y}_l^{-\frac{\theta+1}{\theta}}\right)^{-\theta}\left(\hat{y}_l\hat{W}_l^{-1}\right)^{-\theta}=\left(\frac{E_l}{E_l+\sum_{k\in\mathscr{F}_D(l)}\Xi_{lk}}\right)\hat{y}_l+\sum_{k\in\mathscr{F}_D(l)}\left(\frac{\Xi_{lk}}{E_l+\sum_{k\in\mathscr{F}_D(l)}\Xi_{lk}}\right)\hat{t}_{lk}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{lk,1},\hat{t}_{lk,m}\right)\left(\hat{y}_l\hat{W}_l^{-1}\right)^{-\theta}\left(\hat{l}_k^{\alpha+1}\hat{y}_k^{-\frac{\theta+1}{\theta}}\right)^{-\theta}$$

$$\left(\hat{y}_l^{-\frac{\theta+1}{\theta}}\right)^{-\theta}\left(\hat{y}_l\hat{W}_l^{-1}\right)^{-\theta}=\left(\frac{Y_l}{Y_l+\sum_{k\in\mathscr{F}_D(l)}\Xi_{kl}}\right)\hat{y}_l+\sum_{k\in\mathscr{F}_D(l)}\left(\frac{\Xi_{kl}}{Y_l+\sum_{k\in\mathscr{F}_D(l)}\Xi_{kl}}\right)\hat{t}_{kl}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{kl,1},\hat{t}_{kl,m}\right)\left(\hat{y}_k\hat{l}_k^{\beta-1}\hat{W}^{-1}\right)^{-\theta}\left(\hat{y}_l^{-\frac{\theta+1}{\theta}}\right)^{-\theta}$$

Multiplying both sides by $\hat{W}^{-\theta}$, and applying the definition $\hat{\chi}=\hat{W}^{-\theta}$,

$$\left(\hat{l}_i^{\alpha+1}\hat{y}_i^{-\frac{\theta+1}{\theta}}\right)^{-\theta}\left(\hat{y}_i\hat{l}_i^{\beta-1}\right)^{-\theta}\hat{\chi}^{-1}=\left(\frac{E_i}{E_i+\sum_{k\in\mathscr{F}(i)}\Xi_{ik}}\right)\hat{y}_i+\hat{\chi}^{-1}\sum_{k\in\mathscr{F}_D(i)}\left(\frac{\Xi_{ik}}{E_i+\sum_{k\in\mathscr{F}(i)}\Xi_{ik}}\right)\hat{t}_{ik}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{ik,1},\hat{t}_{ik,m}\right)\left(\hat{y}_i\hat{l}_i^{\beta-1}\right)^{-\theta}\left(\hat{l}_k^{\alpha+1}\hat{y}_k^{-\frac{\theta+1}{\theta}}\right)^{-\theta}+\hat{\chi}^{-1}\sum_{l\in\mathscr{F}_F(i)}\left(\frac{\Xi_{il}}{E_i+\sum_{k\in\mathscr{F}(i)}\Xi_{ik}}\right)\hat{t}_{il}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{ik,1},\hat{t}_{ik,m}\right)\left(\hat{y}_i\hat{l}_i^{\beta-1}\right)^{-\theta}\left(\hat{y}_l^{-\frac{\theta+1}{\theta}}\right)^{-\theta}$$

$$\left(\hat{l}_i^{\alpha+1}\hat{y}_i^{-\frac{\theta+1}{\theta}}\right)^{-\theta}\left(\hat{y}_i\hat{l}_i^{\beta-1}\right)^{-\theta}\hat{\chi}^{-1}=\left(\frac{Y_i}{Y_i+\sum_{k\in\mathscr{F}(i)}\Xi_{ki}}\right)\hat{y}_i+\hat{\chi}^{-1}\sum_{k\in\mathscr{F}_D(i)}\left(\frac{\Xi_{ki}}{Y_i+\sum_{k\in\mathscr{F}(i)}\Xi_{ki}}\right)\hat{t}_{ki}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{ki,1},\hat{t}_{ki,m}\right)\left(\hat{y}_k\hat{l}_k^{\beta-1}\right)^{-\theta}\left(\hat{l}_i^{\alpha+1}\hat{y}_i^{-\frac{\theta+1}{\theta}}\right)^{-\theta}+\sum_{l\in\mathscr{F}_F(i)}\left(\frac{\Xi_{li}}{Y_i+\sum_{k\in\mathscr{F}(i)}\Xi_{ki}}\right)\hat{t}_{li}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{ki,1},\hat{t}_{ki,m}\right)\left(\hat{y}_l\right)^{-\theta}\hat{\chi}_l^{-1}\left(\hat{l}_i^{\alpha+1}\hat{y}_i^{-\frac{\theta+1}{\theta}}\right)^{-\theta}$$

$$\left(\hat{y}_l^{-\frac{\theta+1}{\theta}}\right)^{-\theta}\left(\hat{y}_l\right)^{-\theta}\hat{\chi}_l^{-1}=\left(\frac{E_l}{E_l+\sum_{k\in\mathscr{F}_D(l)}\Xi_{lk}}\right)\hat{y}_l+\sum_{k\in\mathscr{F}_D(l)}\left(\frac{\Xi_{lk}}{E_l+\sum_{k\in\mathscr{F}_D(l)}\Xi_{lk}}\right)\hat{t}_{lk}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{lk,1},\hat{t}_{lk,m}\right)\left(\hat{y}_l\right)^{-\theta}\hat{\chi}_l^{-1}\left(\hat{l}_k^{\alpha+1}\hat{y}_k^{-\frac{\theta+1}{\theta}}\right)^{-\theta}$$

$$\left(\hat{y}_l^{-\frac{\theta+1}{\theta}}\right)^{-\theta}\left(\hat{y}_l\right)^{-\theta}\hat{\chi}_l^{-1}=\left(\frac{Y_l}{Y_l+\sum_{k\in\mathscr{F}_D(l)}\Xi_{kl}}\right)\hat{y}_l+\hat{\chi}^{-1}\sum_{k\in\mathscr{F}_D(l)}\left(\frac{\Xi_{kl}}{Y_l+\sum_{k\in\mathscr{F}_D(l)}\Xi_{kl}}\right)\hat{t}_{kl}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{kl,1},\hat{t}_{kl,m}\right)\left(\hat{y}_k\hat{l}_k^{\beta-1}\right)^{-\theta}\left(\hat{y}_l^{-\frac{\theta+1}{\theta}}\right)^{-\theta}$$

Simplifying,

$$\left(\hat{l}_i^{\alpha+1}\hat{y}_i^{-\frac{\theta+1}{\theta}}\right)^{-\theta}\left(\hat{y}_i\hat{l}_i^{\beta-1}\right)^{-\theta} = \hat{\chi}\left(\frac{E_i}{E_i+\sum_{k\in\mathscr{F}(i)}\Xi_{ik}}\right)\hat{y}_i + \sum_{k\in\mathscr{F}_D(i)}\left(\frac{\Xi_{ik}}{E_i+\sum_{k\in\mathscr{F}(i)}\Xi_{ik}}\right)\hat{t}_{ik}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{ik,1},\hat{t}_{ik,m}\right)\left(\hat{y}_i\hat{l}_i^{\beta-1}\right)^{-\theta}\left(\hat{l}_k^{\alpha+1}\hat{y}_k^{-\frac{\theta+1}{\theta}}\right)^{-\theta} + \sum_{l\in\mathscr{F}_F(i)}\left(\frac{\Xi_{il}}{E_i+\sum_{k\in\mathscr{F}(i)}\Xi_{ik}}\right)\hat{t}_{il}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{ik,1},\hat{t}_{ik,m}\right)\left(\hat{y}_i\hat{l}_i^{\beta-1}\right)^{-\theta}\left(\hat{y}_l^{-\frac{\theta+1}{\theta}}\right)^{-\theta}$$

$$\left(\hat{l}_i^{\alpha+1}\hat{y}_i^{-\frac{\theta+1}{\theta}}\right)^{-\theta}\left(\hat{y}_i\hat{l}_i^{\beta-1}\right)^{-\theta} = \hat{\chi}\left(\frac{Y_i}{Y_i+\sum_{k\in\mathscr{F}(i)}\Xi_{ki}}\right)\hat{y}_i + \sum_{k\in\mathscr{F}_D(i)}\left(\frac{\Xi_{ki}}{Y_i+\sum_{k\in\mathscr{F}(i)}\Xi_{ki}}\right)\hat{t}_{ki}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{ki,1},\hat{t}_{ki,m}\right)\left(\hat{y}_k\hat{l}_k^{\beta-1}\right)^{-\theta}\left(\hat{l}_i^{\alpha+1}\hat{y}_i^{-\frac{\theta+1}{\theta}}\right)^{-\theta} + \sum_{l\in\mathscr{F}_F(i)}\left(\frac{\Xi_{li}}{Y_i+\sum_{k\in\mathscr{F}(i)}\Xi_{ki}}\right)\hat{t}_{li}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{ki,1},\hat{t}_{ki,m}\right)\left(\frac{\hat{\chi}}{\hat{\chi}_l}\right)(\hat{y}_l)^{-\theta}\left(\hat{l}_i^{\alpha+1}\hat{y}_i^{-\frac{\theta+1}{\theta}}\right)^{-\theta}$$

$$\left(\hat{y}_l^{-\frac{\theta+1}{\theta}}\right)^{-\theta}(\hat{y}_l)^{-\theta} = \hat{\chi}_l\left(\frac{E_l}{E_l+\sum_{k\in\mathscr{F}_D(l)}\Xi_{lk}}\right)\hat{y}_l + \sum_{k\in\mathscr{F}_D(l)}\left(\frac{\Xi_{lk}}{E_l+\sum_{k\in\mathscr{F}_D(l)}\Xi_{lk}}\right)\hat{t}_{lk}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{lk,1},\hat{t}_{lk,m}\right)(\hat{y}_l)^{-\theta}\left(\hat{l}_k^{\alpha+1}\hat{y}_k^{-\frac{\theta+1}{\theta}}\right)^{-\theta}$$

$$\left(\hat{y}_l^{-\frac{\theta+1}{\theta}}\right)^{-\theta}(\hat{y}_l)^{-\theta}\hat{\chi}_l^{-1} = \left(\frac{Y_l}{Y_l+\sum_{k\in\mathscr{F}_D(l)}\Xi_{kl}}\right)\hat{y}_l + \hat{\chi}^{-1}\sum_{k\in\mathscr{F}_D(l)}\left(\frac{\Xi_{kl}}{Y_l+\sum_{k\in\mathscr{F}_D(l)}\Xi_{kl}}\right)\hat{t}_{kl}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{kl,1},\hat{t}_{kl,m}\right)\left(\hat{y}_k\hat{l}_k^{\beta-1}\right)^{-\theta}\left(\hat{y}_l^{-\frac{\theta+1}{\theta}}\right)^{-\theta}$$

Simplifying on the LHS and isolating the i specific terms on the second summand on the RHS,

$$\left(\hat{l}_i^{\alpha+1}\hat{y}_i^{-\frac{\theta+1}{\theta}}\right)^{-\theta} = \hat{\chi}\left(\frac{E_i}{E_i+\sum_{k\in\mathscr{F}(i)}\Xi_{ik}}\right)\hat{y}_i\left(\hat{y}_i\hat{l}_i^{\beta-1}\right)^{\theta} + \sum_{k\in\mathscr{F}_D(i)}\left(\frac{\Xi_{ik}}{E_i+\sum_{k\in\mathscr{F}(i)}\Xi_{ik}}\right)\hat{t}_{ik}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{ik,1},\hat{t}_{ik,m}\right)\left(\hat{l}_k^{\alpha+1}\hat{y}_k^{-\frac{\theta+1}{\theta}}\right)^{-\theta} + \sum_{l\in\mathscr{F}_F(i)}\left(\frac{\Xi_{il}}{E_i+\sum_{k\in\mathscr{F}(i)}\Xi_{ik}}\right)\hat{t}_{il}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{ik,1},\hat{t}_{ik,m}\right)\left(\hat{y}_l^{-\frac{\theta+1}{\theta}}\right)^{-\theta}$$

$$\left(\hat{y}_i\hat{l}_i^{\beta-1}\right)^{-\theta} = \hat{\chi}\left(\frac{Y_i}{Y_i+\sum_{k\in\mathscr{F}(i)}\Xi_{ki}}\right)\hat{y}_i\left(\hat{l}_i^{\alpha+1}\hat{y}_i^{-\frac{\theta+1}{\theta}}\right)^{\theta} + \sum_{k\in\mathscr{F}_D(i)}\left(\frac{\Xi_{ki}}{Y_i+\sum_{k\in\mathscr{F}(i)}\Xi_{ki}}\right)\hat{t}_{ki}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{ki,1},\hat{t}_{ki,m}\right)\left(\hat{y}_k\hat{l}_k^{\beta-1}\right)^{-\theta} + \sum_{l\in\mathscr{F}_F(i)}\left(\frac{\Xi_{li}}{Y_i+\sum_{k\in\mathscr{F}(i)}\Xi_{ki}}\right)\hat{t}_{li}^{-\theta}\left(\hat{\Pi},\hat{P},\hat{t}_{ki,1},\hat{t}_{ki,m}\right)\left(\frac{\hat{\chi}}{\hat{\chi}_l}\right)(\hat{y}_l)^{-\theta}$$

Similarly for foreign locations,

$$\left(\hat{y}_l^{-\frac{\theta+1}{\theta}}\right)^{-\theta} = \hat{\chi}_l \left(\frac{E_l}{E_l + \sum_{k\in\mathscr{F}_D(l)} \Xi_{lk}}\right) \hat{y}_l \left(\hat{y}_l\right)^{\theta} + \sum_{k\in\mathscr{F}_D(l)} \left(\frac{\Xi_{lk}}{E_l + \sum_{k\in\mathscr{F}_D(l)} \Xi_{lk}}\right) \hat{t}_{lk}^{-\theta} \left(\hat{\Pi}, \hat{P}, \hat{t}_{lk,1}, \hat{t}_{lk,m}\right) \left(\hat{l}_k^{\alpha+1} \hat{y}_k^{-\frac{\theta+1}{\theta}}\right)^{-\theta}$$

$$\left(\hat{y}_l \hat{W}_l^{-1}\right)^{-\theta} \hat{\chi}_l^{-1} = \left(\frac{Y_l}{Y_l + \sum_{k\in\mathscr{F}_D(l)} \Xi_{kl}}\right) \hat{y}_l \left(\hat{y}_l^{-\frac{\theta+1}{\theta}}\right)^{\theta} + \hat{\chi}^{-1} \sum_{k\in\mathscr{F}_D(l)} \left(\frac{\Xi_{kl}}{Y_l + \sum_{k\in\mathscr{F}_D(l)} \Xi_{kl}}\right) \hat{t}_{kl}^{-\theta} \left(\hat{\Pi}, \hat{P}, \hat{t}_{kl,1}, \hat{t}_{kl,m}\right) \left(\hat{y}_k \hat{l}_k^{\beta-1}\right)^{-\theta}$$

This system of equation pins down the equilibrium with foreign locations:

$$\hat{y}_i^{1+\theta} \hat{l}_i^{-\theta(\alpha+1)} = \hat{\chi} \left(\frac{E_i}{E_i + \sum_{k\in\mathscr{F}(i)} \Xi_{ik}}\right) \hat{y}_i^{1+\theta} \hat{l}_i^{\theta(\beta-1)} + \sum_{k\in\mathscr{F}_D(i)} \left(\frac{\Xi_{ik}}{E_i + \sum_{k\in\mathscr{F}(i)} \Xi_{ik}}\right) \hat{t}_{ik}^{-\theta} \left(\hat{\Pi}, \hat{P}, \hat{t}_{ik,1}, \hat{t}_{ik,m}\right) \hat{y}_k^{1+\theta} \hat{l}_k^{-\theta(1+\alpha)} + \sum_{l\in\mathscr{F}_F(i)} \left(\frac{\Xi_{il}}{E_i + \sum_{k\in\mathscr{F}(i)} \Xi_{ik}}\right) \hat{t}_{il}^{-\theta} \left(\hat{\Pi}, \hat{P}, \hat{t}_{ik,1}, \hat{t}_{ik,m}\right) \hat{y}_l^{1+\theta}$$

$$\hat{y}_i^{-\theta} \hat{l}_i^{\theta(1-\beta)} = \hat{\chi} \left(\frac{Y_i}{Y_i + \sum_{k\in\mathscr{F}(i)} \Xi_{ki}}\right) \hat{y}_i^{-\theta} \hat{l}_i^{\theta(\alpha+1)} + \sum_{k\in\mathscr{F}_D(i)} \left(\frac{\Xi_{ki}}{Y_i + \sum_{k\in\mathscr{F}(i)} \Xi_{ki}}\right) \hat{t}_{ki}^{-\theta} \left(\hat{\Pi}, \hat{P}, \hat{t}_{ki,1}, \hat{t}_{ki,m}\right) \hat{y}_k^{-\theta} \hat{l}_k^{\theta(1-\beta)} + \hat{\chi} \sum_{l\in\mathscr{F}_F(i)} \left(\frac{\Xi_{li}}{Y_i + \sum_{k\in\mathscr{F}(i)} \Xi_{ki}}\right) \hat{t}_{li}^{-\theta} \left(\hat{\Pi}, \hat{P}, \hat{t}_{ki,1}, \hat{t}_{ki,m}\right) \hat{\chi}_l^{-1} \hat{y}_l^{-\theta}$$

$$\hat{y}_l^{1+\theta} = \left(\frac{E_l}{E_l + \sum_{k\in\mathscr{F}_D(l)} \Xi_{lk}}\right) \hat{y}_l^{1+\theta} \hat{\chi}_l + \sum_{k\in\mathscr{F}_D(l)} \left(\frac{\Xi_{lk}}{E_l + \sum_{k\in\mathscr{F}_D(l)} \Xi_{lk}}\right) \hat{t}_{lk}^{-\theta} \left(\hat{\Pi}, \hat{P}, \hat{t}_{lk,1}, \hat{t}_{lk,m}\right) \hat{y}_k^{1+\theta} \hat{l}_k^{-\theta(1+\alpha)}$$

$$\hat{y}_l^{-\theta} \hat{\chi}_l^{-1} = \left(\frac{Y_l}{Y_l + \sum_{k\in\mathscr{F}_D(l)} \Xi_{kl}}\right) \hat{y}_l^{-\theta} + \hat{\chi}^{-1} \sum_{k\in\mathscr{F}_D(l)} \left(\frac{\Xi_{kl}}{Y_l + \sum_{k\in\mathscr{F}_D(l)} \Xi_{kl}}\right) \hat{t}_{kl}^{-\theta} \left(\hat{\Pi}, \hat{P}, \hat{t}_{kl,1}, \hat{t}_{kl,m}\right) \hat{y}_k^{-\theta} \hat{l}_k^{\theta(1-\beta)}$$

A.13 Derivations for Section 6.2: Counterfactuals with Fixed-Routing

This subsection gives a self-contained description of how a single counterfactual is solved under fixed routing and how the associated welfare change is recovered. The maintained assumptions are: (i) the baseline OD–segment incidence kernel is held fixed throughout the counterfactual; (ii) trade demand is CES with elasticity $\sigma > 1$; and (iii) labor is perfectly mobile across locations, so equilibrium welfare is equalized.

Fixed routing. Consider one counterfactual vector of exogenous segment-level free-flow shocks $\{\hat{\bar{t}}_{k\ell,m}\}$. Let $\rho_{ij}^{k\ell,m} \geq 0$ denote the baseline OD–segment incidence weights, normalized so that

$$\sum_{(k,\ell),m} \rho_{ij}^{k\ell,m} = 1 \qquad \text{for each } (i,j) \text{ with } X_{ij}^0 > 0.$$

Holding these weights fixed is the no-rerouting restriction. Under fixed routing, we define the delivered trade-cost index between origin i and destination j by

$$\tau_{ij} := \prod_{(k,\ell),m} t_{k\ell,m}^{\rho_{ij}^{k\ell,m}}, \tag{A.88}$$

so that

$$\ln \hat{\tau}_{ij} = \sum_{(k,\ell),m} \rho_{ij}^{k\ell,m} \ln \hat{t}_{k\ell,m}. \tag{A.89}$$

Equation (A.88) is the maintained fixed-routing unit-cost index; with this definition, $\rho_{ij}^{k\ell,m}$ is exactly the marginal incidence elasticity of bilateral trade costs with respect to segment costs.

If congestion is absent, then $\hat{t}_{k\ell,m} = \hat{\bar{t}}_{k\ell,m}$ and $\hat{\tau}_{ij}$ is obtained directly from (A.89). If congestion is present, current OD trade hats imply segment traffic

$$\Xi'_{k\ell,m} = \sum_{i=1}^{N}\sum_{j=1}^{N} \rho_{ij}^{k\ell,m} X_{ij}^0 \hat{X}_{ij}, \qquad \hat{\Xi}_{k\ell,m} := \frac{\Xi'_{k\ell,m}}{\Xi^0_{k\ell,m}}, \tag{A.90}$$

and segment costs satisfy

$$\hat{t}_{k\ell,m} = \hat{\bar{t}}_{k\ell,m} (\hat{\Xi}_{k\ell,m})^{\lambda_m}. \tag{A.91}$$

Hence, under fixed routing, the transport block maps current OD trade hats into bilateral trade-cost hats through

$$\hat{X} \longrightarrow \hat{\Xi} \longrightarrow \hat{t} \longrightarrow \hat{\tau}.$$

When the observed baseline traffic $\Xi^0_{k\ell,m}$ is not exactly reproduced by the baseline decomposition $\sum_{i,j} \rho_{ij}^{k\ell,m} X_{ij}^0$, the denominator in (A.90) should be replaced by the model-implied baseline traffic so that $\hat{\Xi}_{k\ell,m} = 1$ when $\hat{X}_{ij} = 1$ for all (i,j).

Counterfactual equilibrium. Let $\theta := \sigma - 1 > 0$. In location i, the producer price is

$$p_i = \frac{w_i}{A_i},$$

where w_i is the wage and A_i is productivity. Consumers in location j have indirect utility

$$U_j = \frac{u_j w_j}{P_j},$$

where u_j is amenity and P_j is the CES price index. Perfect mobility implies welfare equalization,

$$U_j = W \qquad \forall j,$$

for some common welfare level W.

Bilateral trade flows satisfy the CES demand system

$$X_{ij} = \left(\frac{p_i \tau_{ij}}{P_j}\right)^{1-\sigma} E_j, \qquad E_j = w_j L_j. \tag{A.92}$$

Using $p_i = w_i/A_i$ and $P_j = u_j w_j/W$, equation (A.92) becomes

$$X_{ij} = W^{1-\sigma}\left(\frac{A_i u_j}{\tau_{ij}}\right)^{\sigma-1} w_i^{1-\sigma} w_j^{\sigma} L_j. \tag{A.93}$$

Define

$$\chi := W^{1-\sigma} = W^{-\theta}.$$

Then (A.93) can be written as

$$X_{ij} = \chi\left(\frac{A_i u_j}{\tau_{ij}}\right)^{\theta} w_i^{-\theta} w_j^{\theta+1} L_j,$$

and taking hats gives

$$\hat{X}_{ij} = \hat{\chi}\left(\frac{\hat{A}_i \hat{u}_j}{\hat{\tau}_{ij}}\right)^{\theta} \hat{w}_i^{-\theta} \hat{w}_j^{\theta+1} \hat{L}_j. \tag{A.94}$$

Goods-market clearing requires local labor income to equal total sales:

$$w_i L_i = \sum_{j=1}^{N} X_{ij}.$$

Dividing by the corresponding baseline identity $w_i^0 L_i^0 = \sum_j X_{ij}^0$ yields

$$\hat{w}_i \hat{L}_i = \sum_{j=1}^{N} s_{ij}^{X,0} \hat{X}_{ij}, \qquad s_{ij}^{X,0} := \frac{X_{ij}^0}{w_i^0 L_i^0}. \tag{A.95}$$

The CES price index satisfies

$$P_i^{-\theta} = \sum_{j=1}^{N} (p_j \tau_{ji})^{-\theta}.$$

Its exact hat form is therefore

$$\hat{P}_i^{-\theta} = \sum_{j=1}^{N} \pi_{ji}^0 (\hat{p}_j \hat{\tau}_{ji})^{-\theta} = \sum_{j=1}^{N} \pi_{ji}^0 \left(\frac{\hat{A}_j}{\hat{w}_j \hat{\tau}_{ji}}\right)^{\theta}, \tag{A.96}$$

where

$$\pi_{ji}^0 := \frac{X_{ji}^0}{E_i^0}, \qquad E_i^0 = w_i^0 L_i^0, \qquad \hat{p}_j = \frac{\hat{w}_j}{\hat{A}_j}.$$

Because mobility implies $P_i = u_i w_i / W$, taking hats gives

$$\hat{P}_i = \frac{\hat{u}_i \hat{w}_i}{\hat{W}}.$$

Rearranging yields

$$\hat{w}_i^{-\theta} = \hat{W}^{-\theta} \hat{u}_i^{\theta} \hat{P}_i^{-\theta}.$$

Substituting (A.96) and using $\hat{\chi} = \hat{W}^{-\theta}$ gives the price-index/mobility equation

$$\hat{w}_i^{-\theta} = \hat{\chi}\, \hat{u}_i^{\theta} \sum_{j=1}^{N} \pi_{ji}^0 \left(\frac{\hat{A}_j}{\hat{w}_j \hat{\tau}_{ji}}\right)^{\theta}. \tag{A.97}$$

To close the system, suppose productivity and amenities satisfy

$$A_i = \bar{A}_i L_i^{\alpha}, \qquad u_i = \bar{u}_i L_i^{\beta}.$$

If the counterfactual changes only trade costs and leaves $(\bar{A}_i, \bar{u}_i)$ fixed, then

$$\hat{A}_i = \hat{L}_i^{\alpha}, \qquad \hat{u}_i = \hat{L}_i^{\beta}. \tag{A.98}$$

More generally, with exogenous shifts in fundamentals one would use $\hat{A}_i = \hat{\bar{A}}_i \hat{L}_i^{\alpha}$ and $\hat{u}_i = \hat{\bar{u}}_i \hat{L}_i^{\beta}$. Finally, impose the population normalization

$$\sum_{i=1}^{N} \lambda_i \hat{L}_i = 1, \qquad \lambda_i := \frac{L_i^0}{\sum_k L_k^0}. \tag{A.99}$$

Equations (A.94), (A.95), (A.97), (A.98), and (A.99), together with the fixed-routing transport block (A.89)–(A.91), characterize the counterfactual equilibrium.

Solving for $\hat{\chi}$ and recovering welfare. For the given counterfactual $\{\hat{\bar{t}}_{k\ell,m}\}$, the routing kernel $\rho_{ij}^{k\ell,m}$ is held fixed throughout. Given a trial value of $\hat{\chi}$, we solve the inner equilibrium for the implied objects $(\hat{w}, \hat{L}, \hat{X}, \hat{\tau})$. When congestion is absent, $\hat{\tau}$ is computed once from (A.89) and the inner problem reduces to solving (A.94), (A.95), (A.97), and (A.98). When congestion is present, we iterate between the trade block (A.94), (A.95), (A.97), (A.98) and the transport block (A.90)–(A.91) until $(\hat{X}, \hat{\tau})$ converge.

This inner solve delivers $\hat{L}(\hat{\chi})$ and therefore the scalar residual

$$r(\hat{\chi}) := \sum_{i=1}^{N} \lambda_i \hat{L}_i(\hat{\chi}) - 1.$$

The outer problem is one-dimensional: we update $\hat{\chi}$ with a scalar root finder until $r(\hat{\chi}) = 0$. Once the root is found, the common welfare change follows immediately from the definition of χ:

$$\hat{\chi} = \hat{W}^{-\theta} \qquad \Longleftrightarrow \qquad \hat{W} = \hat{\chi}^{-1/\theta}. \tag{A.100}$$

Equivalently, welfare can be recovered from any location as

$$\hat{W}_i = \frac{\hat{u}_i \hat{w}_i}{\hat{P}_i}, \tag{A.101}$$

and, at the solution, $\hat{W}_i$ is constant across i and coincides with $\hat{\chi}^{-1/\theta}$ up to numerical error.

B Data and Additional Results

B.1 Data Appendix

Figure A.1 captures the value share of goods that are transported within the United States by transport mode and distance.

B.1.1 Rail Data

Our confidential carload waybills data comes from the Surface Transportation Board. This is a stratified sample of carload waybills for all U.S. rail traffic submitted by those rail carriers terminating 4,500 or more revenue carloads annually, covering 48 states (except Alaska and Hawaii). The waybills data report the origin rail station, destination rail station, and the interchange stations in between that these freight cargo are transported through. The rich geographical information in this confidential data set allows us to study the routing of these commodities through the railroad network. Additionally, this data set also contains commodity-specific information including number of car loads, weight, whether it is a domestic or international shipment, and its inter-modality—primarily indicating if the freight movement involved other transport modes, which is almost entirely containers. The intermodal container rail traffic is the fastest growing segment of rail traffic, having grown by more than 5 times between 1984 and 2019 (Figure A.2).

B.1.2 Road Data

We follow Allen and Arkolakis (2022) for the construction of road traffic flows data. We summarize their procedure in three steps: First, they create a sparse graph representation of the underlying road network by collapsing the high-dimensional geo-spatial information contained in the original shapefiles and only preserving nodes that are either endpoints or intersections. Furthermore, core-based statistical areas (CBSAs) are represented by a singular node along the network. Their resulting graph consists of 228 nodes and 704 edges. Second, they construct a weighted graph by including traffic data. To do so they obtain the average annual daily traffic from the 2012 Highway Performance Monitoring System (HPMS) dataset by the Federal Highway Administration and allocate it to individual links by constructing a length-weighted average of the annual daily traffic.

B.1.3 Ports

While we include the top container ports in the U.S. in our analysis, we merge some of these ports into the one polygon due to the U.S.ACE-provided polygons either overlapping or being very close to each other. Specifically, Tacoma + Seattle are merged into a single polygon due to their port alliance (NWSA). Additionally, Tacoma and Seattle share a port alliance. Tampa + Manatee were merged do to overlap issues between the rather large Tampa region and the manually applied Manatee port statistical area. Los Angeles and Long Beach are also merged due to proximity. Lastly, we include Chester in the Philadelphia PSA because the U.S.ACE-provided polygon area includes Chester. Between the 28 coastal and inland ports, we capture 98 percent of total U.S. container volumes.

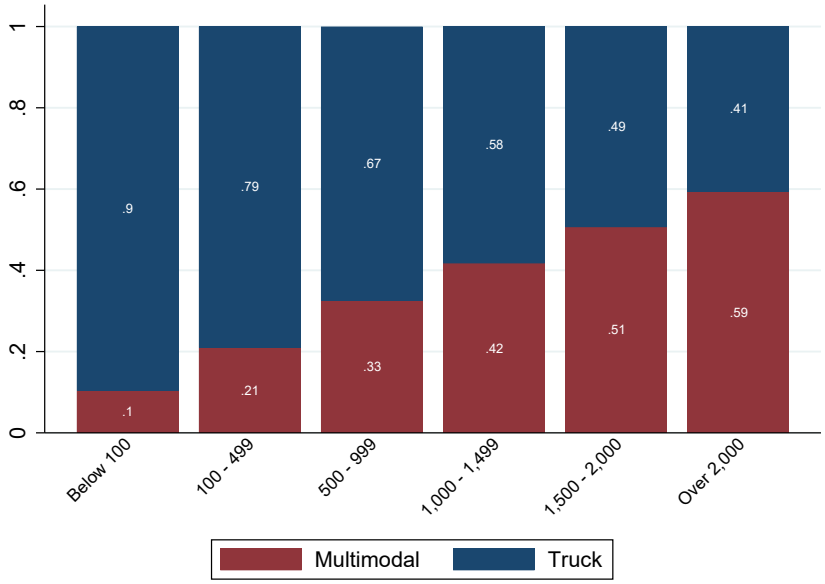
B.1.4 Automatic Identification System (AIS) Vessel Traffic Data

We utilize automatic identification system (AIS) vessel traffic data from Marine Cadastre, a joint initiative between the Bureau of Ocean Energy Management and the National Oceanic and Atmospheric Administration. Here we highlight two examples to show how we capture these ships and the time they spend at a port. Panel (A) Figure A.3 shows the path of containership CMA CGM Christophe Colomb as it enters the Port of Los Angeles (LA) on May 2, 2022. It is a containership with a cargo capacity of 86,100 tons (13,800 twenty-foot equivalent unit containers (TEUs)) and is operated by container shipping company CMA CGM. Panel (B) Figure A.3 shows the path of containership Guthorm Maersk entering and leaving the Port of Newark. Guthorm Maersk is a containership with a cargo capacity of 57,000 tons (11,000 TEUs) and is operated by container shipping company Maersk. The ship path entering the port is highlighted in the figure and the redder color indicates slower speed. The darker region of both figures indicate the port polygon for both ports as defined by the U.S. Army Corps of Engineers.

B.1.5 Matching Rail Traffic Data to Metropolitan Statistical Area (MSAs)

Using 1999 Metropolitan Statistical Area (MSA) polygons from the Census Bureau, we match the rail stations from the Waybill data to the 228 MSAs in Duranton and Turner (2011). By rail destination, we observe 224 MSAs and by origin we

Figure A.1. U.S. Transport Mode Value Shares by Distance



Notes: This figure plots the observed value share of cargo transported by different modes across various distances in 2018. Multimodal indicates cargo movement that involves more than one mode, including truck, rail, and waterways. Source: Freight Analysis Framework, U.S. Department of Transportation, and authors' calculations.

observe 223 MSAs. Since some of these unobserved MSAs overlap, we have 7 MSAs. We conduct our analysis on the remaining 221 MSAs, so that both the origin and destination results are comparable. The 7 MSAs that we do not have rail traffic for Daytona Beach FL, Fort Myers-Cape Coral FL, Fort Walton Beach FL, New Haven-Bridgeport-Stamford-Waterbury-Danbury CT, Providence-Warwick-Pawtucket RI, Punta Gorda FL, and Santa Fe NM.

To match the MSA-level truck vehicle-kilometers traveled (VKT) measure in Duranton and Turner (2011), we calculate the rail equivalent in two ways. First, we utilize the number of rail carloads, transported in and out of MSAs, multiplied by the weighted average of their distance traveled. We call this railcar-kilometers-traveled (rail VKT) and distinguish by destination for rail shipments transported into these MSAs, and by origin for shipments transported out of these MSAs. We also observe the weight of these rail shipments and can calculate rail weight VKT using the same method outlined previously.

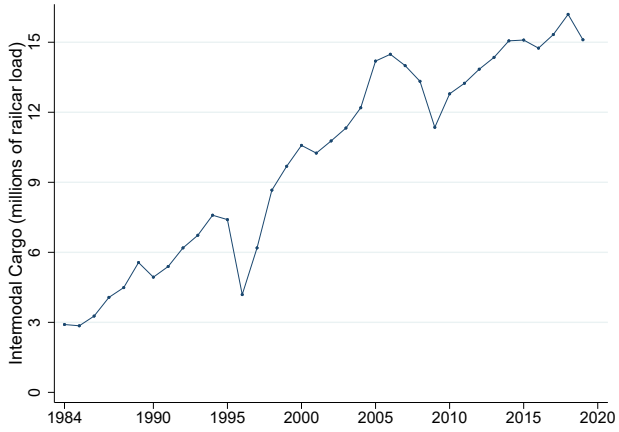
B.1.6 Rail Dwell Times Data

We obtain weekly rail station dwell times from the Surface Transportation Board (STB). Railroads provide the STB with the average time a railcar resides at a station, measured in hours, for their 10 largest stations in terms of railcars processed. This dwell time measure excludes cars on through trains—trains that travels without stops en route. Since this dataset only captures a subset of all rail stations (albeit the largest ones), we match the ports in the previous section to their local rail stations. We do this by expanding the port polygon areas in 50km intervals. The rail stations that are captured in the buffer areas of their closest port is be considered a rail station in the vicinity of this port and is likely to service traffic to and from the port. Due to their proximity, The ports of Los Angeles and Long Beach and combined into one port for this exercise. We use a buffer area of 150km which captures 7 ports and 12 rail stations. We test the robustness of this buffer area by increasing the interval in our analysis to 200km where we capture 8 ports and 14 rail stations. Further increases to this interval result in more muted responses of rail station dwell times to port traffic, as these rail stations are much further away.

Additionally, the rail dwell times dataset is reported at the weekly level. In order to match this to our daily port traffic measure for analysis, we aggregate our port traffic measure up to the weekly level. We start our week on a Monday since we observe in our data that most ships tend to enter a port on Mondays.

Summary Statistics Figure A.4 plots the average of rail station dwell times from June 2015 to December 2021. The average dwell time over this period is around 25.5 hours per station with a standard deviation of 2.5 hours. However, there is also a large decrease in dwell times around the start of the pandemic followed up a steep increase afterwards.

Figure A.2. Intermodal Container Rail Traffic, 1984-2019



Notes: This figure plots the observed levels of intermodal rail cargo from 1984 to 2019. Source: Confidential Carload Waybill, Surface Transportation Board, and authors' calculations.

B.2 Regression design for Modal Substitution (Subsection 4.1)

We are interested in deriving a regression that studies the impact of changes in the transport network on the modal mix in order to identify the elasticity of substitution between modes. In Subsection 4.1 we examine the results of a regression that examines location (i.e. MSA) specific outcomes as a function of local changes in the primary network. Conceptually, the regression can be seen understood to be of the following form:

$$d \ln Y_k = \alpha + \beta \times d \ln \bar{t}_{k,1} + \epsilon_{kk,1} \tag{B.1}$$

where Y_k refers to a local outcome (primary network traffic, or the ratio between primary and non-primary traffic) in a MSA located at node k and $d \ln \bar{t}_k$ refers to changes in transportation cost of the localized primary network.In what follows we motivate this regression by deriving a structural interpretation. In what follows, we will make an additional assumption that there exists some localized primary network fully contained within the MSA that any unimodal route originating or terminating in k needs to transition through before accessing the national primary road network. Let this localized network be represented by the transportation cost $\bar{t}_k$. In order to derive the regression we start by examining modal traffic (Equation (3.9)), which we restate here for convenience,

$$\Xi_{kl,m} = t_{kl,m}^{-\eta} \times t_{kl}^{\eta-\theta} \times P_k^{-\theta} \times \Pi_l^{-\theta}$$

where we can characterize total outgoing traffic on the primary and non-primary network as,

$$\Xi_{kk,1} = P_k^{-\theta} \times t_{k,1}^{-\eta} \times \sum_{l \in \mathscr{F}_1(k)} t_{kl,1}^{-\eta} \times t_{kl}^{\eta-\theta} \times \Pi_l^{-\theta}$$

$$\Xi_{kk,m} = P_k^{-\theta} \times s_{kk,m}^{-\eta} \times \sum_{l \in \mathscr{F}_m(k)} \left(\tau_{kl,m} s_{ll,m}\right)^{-\eta} \times t_{kl}^{\eta-\theta} \times \Pi_l^{-\theta}$$

Totally differentiating with respect to the change in local transport cost on the primary network,

$$d \ln \Xi_{kk,1} = -\theta d \ln P_k - \eta d \ln t_{k,1} + d \ln \left(\sum_{l \in \mathscr{F}_1(k)} t_{kl,1}^{-\eta} \times t_{kl}^{\eta-\theta} \times \Pi_l^{-\theta} \right) \tag{B.2}$$

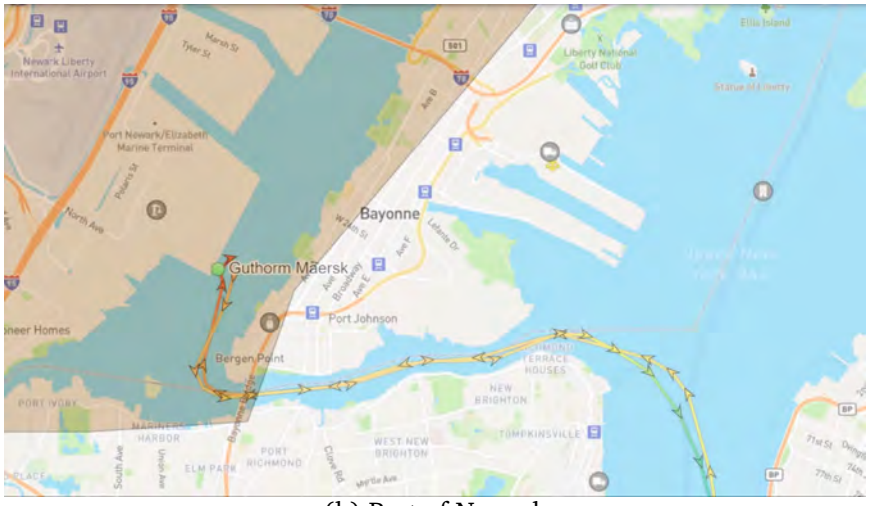
where this capture the impact of highway changes on highway traffic, plus market access changes

$$d \ln \Xi_{kk,m} = -\theta d \ln P_k + d \ln \left(\sum_{l \in \mathscr{F}_m(k)} \left(s_{kk,m} \tau_{kl,m} s_{ll,m}\right)^{-\eta} \times t_{kl}^{\eta-\theta} \times \Pi_l^{-\theta} \right)$$

Figure A.3. Illustration of AIS Mooring Paths



(a) Port of Los Angeles



(b) Port of Newark

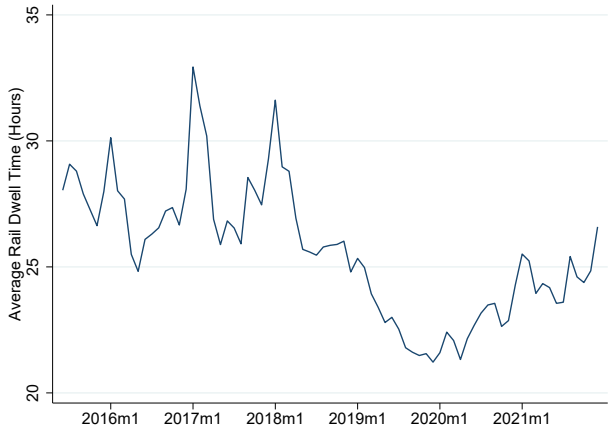
Notes: Panel (a) shows the containership CMA CGM Christophe Colomb at the Port of Los Angeles while Panel (b) shows the containership Guthorm Maersk at the Port of Newark. The path of each ship to and from the port shows its exact travel path. The darker regions at each port shows the port polygons as defined by the U.S. Army Corps of Engineers.

which expresses changes in modal traffic as a function of the local price index ($d \ln P_k$), changes in local transport cost, either for the primary mode ($d \ln t_{k,1}$), and a final summation term that captures changes in mode-specific market access. While we could in principle regress directly on primary mode traffic as in Equation (B.2), the obvious challenge would be the confounding impact of changes in the local price index which will be affected by the local network changes. Instead, we propose to form the ratio between primary and non-primary modal traffic to difference out the local price index. Forming this ratio and taking the total differential, we obtain,

$$d \ln \frac{\Xi_{kk,1}}{\Xi_{kk,m}} = -\eta d \ln t_{k,1} + d \ln \left(\sum_{l \in \mathcal{F}_1(k)} \frac{t_{kl,1}^{-\eta} \times t_{kl}^{\eta-\theta} \times \Pi_l^{-\theta}}{\sum_{l \in \mathcal{F}_m(k)} \left(s_{kk,m} \tau_{kl,m} s_{ll,m} \right)^{-\eta} \times t_{kl}^{\eta-\theta} \times \Pi_l^{-\theta}} \right) \tag{B.3}$$

where instead now the only confounder is the final summation term that summarizes changes in the relative mode-specific market access. Lining this up with the initial motivating regression design (Equation (B.1)), we can see that the coefficient identifies the modal elasticity of substitution ($\beta = \eta$), while the identifying restriction is that the instrument needs to be orthogonal to changes in the relative mode-specific market access.

Figure A.4. Rail Station Dwell Times



Notes: This figure plots the average time a railcar spends at a rail station from June 2015 to July 2022.

B.3 Elasticity of Modal Substitution: Additional Results

Since we have a slightly smaller set of MSAs due to the matching process between MSAs and our rail traffic data (see Section B.1.5 for more information), we first show that we are able to replicate the truck traffic use results from Duranton and Turner (2011) in Table A.2 (first stage results are in Table A.3). Table A.4 presents the rail traffic use results while Table A.5 presents the alternative weight-based measure of the relative truck to rail traffic use. Table A.6 presents additional robustness checks on our baseline results.

Table. A.2. Elasticity of Truck Traffic Use with respect to Road Infrastructure Improvements

	(1)	(2)	(3)	(4)	(5)
Truck Traffic Use	OLS	OLS	IV	IV	IV
Interstate Highway Lane KM	1.606	1.616	1.746	2.083	2.099
	(0.328)	(0.338)	(0.427)	(0.483)	(0.530)
Population		0.967	-0.278	-0.615	-0.484
		(0.550)	(0.303)	(0.376)	(0.393)
Geography				✓	✓
Census Divisions				✓	✓
Socioeconomic Characteristics		✓			✓
MSA FE	✓	✓			
Year FE	✓	✓	✓	✓	✓
Observations	663	663	663	663	663
R-squared	0.77	0.78	0.49	0.49	0.51
KP F-stat			13.48	10.08	10.02

Notes: Robust standard errors clustered by MSAs in parentheses. All variables are in logs. Instruments are 1835 exploration routes, 1898 railroad route kilometers, and 1947 planned interstate highways. 663 observations corresponding to 221 MSAs for each regression. See Table A.3 for first-stage regressions. The results in this table are produced using data from Duranton and Turner (2011) that is matched to rail traffic data.

Table. A.3. First Stage

	(1)	(2)	(3)
1898 Railroads	0.0879	0.0939	0.119
	(0.0460)	(0.0499)	(0.0474)
1947 Planned Interstates	0.156	0.127	0.114
	(0.0332)	(0.0322)	(0.0284)
1835 Exploration Routes	0.0249	0.0268	0.0222
	(0.0117)	(0.0124)	(0.0122)
Population	0.516	0.599	0.545
	(0.0393)	(0.0481)	(0.0597)
Geography		✓	✓
Census Divisions		✓	✓
Socioeconomic Characteristics			✓
Year FE	✓	✓	✓
Observations	663	663	663
KP F-stat	13.48	10.08	10.02

***Notes*:** See Equation (4.2) for more details. Robust standard errors clustered by MSAs in parentheses. All variables are in logs. Instruments are 1835 exploration routes, 1898 railroad route kilometers, and 1947 planned interstate highways. 663 observations corresponding to 221 MSAs for each regression. The results in this table are produced using data from Duranton and Turner (2011) that is matched to rail traffic data.

Table. A.4. Elasticity of Rail Traffic Use with respect to Road Infrastructure Improvements

	(1)	(2)	(3)	(4)	(5)
Rail Traffic Use	OLS	OLS	IV	IV	IV
Interstate Highway Lane KM	-0.103	-0.0993	0.434	0.254	0.401
	(0.173)	(0.175)	(0.314)	(0.337)	(0.315)
Population		0.346	0.695	0.878	0.757
		(0.299)	(0.245)	(0.286)	(0.273)
Geography				✓	✓
Census Divisions				✓	✓
Socioeconomic Characteristics		✓			✓
MSA FE	✓	✓			
Year FE	✓	✓	✓	✓	✓
Observations	663	663	663	663	663
R-squared	0.94	0.94	0.39	0.55	0.57
KP F-stat			13.48	10.08	10.02

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered by MSAs in parentheses. All variables are in logs. Rail traffic use, measured in railcar-kilometers, is constructed using confidential rail waybill data. Truck traffic use (in vehicle-kilometers) and other variables are from Duranton and Turner (2011). Instruments are 1835 exploration routes, 1898 railroad route kilometers, and 1947 planned interstate highways. 663 observations corresponding to 221 MSAs for each regression. See Table A.3 for first stage regressions.

Table. A.5. Elasticity of Modal Substitution using Weight-Based Measure: Ratio of Rail to Truck Traffic Use with respect to Road Infrastructure Improvements

	(1)	(2)	(3)	(4)	(5)
	OLS	OLS	IV	IV	IV
Interstate Highway Lane KM	-1.473	-1.472	-0.930	-1.373	-1.203
	(0.171)	(0.172)	(0.392)	(0.403)	(0.382)
Population		-0.101	0.524	1.012	0.774
		(0.308)	(0.297)	(0.338)	(0.316)
Geography				✓	✓
Census Divisions				✓	✓
Socioeconomic Characteristics		✓			✓
MSA FE	✓	✓			
Year FE	✓	✓	✓	✓	✓
Observations	658	658	658	658	658
R-squared	0.89	0.89	-0.03	0.23	0.28
KP F-stat			14.48	10.76	10.04

Notes: Robust standard errors clustered by MSAs in parentheses. All variables are in logs. Rail traffic use, measured in rail weight-kilometers, is constructed using confidential rail waybill data. Truck traffic use and all other variables are from Duranton and Turner (2011). Instruments are 1835 exploration routes, 1898 railroad route kilometers, and 1947 planned interstate highways. 663 observations corresponding to 221 MSAs for each regression. See Table A.3 for first stage regressions.

Table. A.6. Elasticity of Modal Substitution: Robustness Checks

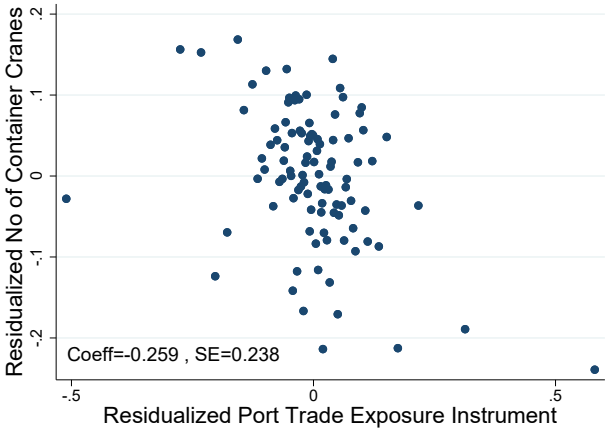
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rail to Truck Traffic Use	IV	IV	IV	IV	IV	IV	IV
Interstate Highway Lane KM	-1.099	-0.999	-1.220	-1.019	-1.283	-1.593	-1.703
	(0.364)	(0.405)	(0.444)	(0.430)	(0.462)	(0.528)	(0.543)
Population	0.891	1.145	1.000	0.973	0.789	1.267	1.155
	(0.306)	(0.336)	(0.367)	(0.352)	(0.376)	(0.434)	(0.443)
Geography	✓	✓	✓	✓	✓	✓	✓
Census Divisions	✓	✓	✓	✓	✓	✓	✓
Socioeconomic Characteristics	✓	✓	✓	✓	✓	✓	✓
MSA FE							
Year FE	✓	✓	✓	✓	✓	✓	✓
Rail Measure	Carload	Carload	Carload	Weight	Weight	Carload	Weight
Without 1898 Rail IV						✓	✓
Data	Total	Incoming	Outgoing	Incoming	Outgoing	Total	Total
Observations	658	658	658	658	658	658	658
R-squared	0.27	0.24	0.31	0.21	0.34	0.20	0.20
KP F-stat	10.04	10.04	10.04	10.04	10.04	14.25	14.25

Notes: Robust standard errors clustered by MSAs in parentheses. All variables are in logs. Rail traffic use, measured in rail weight-kilometers, is constructed using confidential rail waybill data. Truck traffic use (in vehicle-kilometers) and other variables are from Duranton and Turner (2011). Instruments are 1835 exploration routes, 1898 railroad route kilometers, and 1947 planned interstate highways. These observations correspond to 221 MSAs for each regression. See Table A.3 for first stage regressions. Column (1) reproduces the baseline results from Column (5) in Table 1 for comparison. Column (2) restricts the results from Column (1) to just incoming rail traffic use while Column (3) restricts the results to outgoing rail traffic use. For detailed results for weight-based measure of total rail traffic use, see Table A.5. Column (4) uses this weight-based measure of rail traffic use but restricts the results to just incoming traffic while Column (5) restricts the results to outgoing traffic. Columns (6) and (7) reruns the 2SLS regression without the 1898 railroad instrument, for the carload- and weight-based rail traffic use measures respectively.

B.4 Estimation of Intermodal Terminal Congestion: Additional Results

Figure A.5 presents a scatter plot between our residualized port trade exposure instrument and a proxy for unobserved determinants of ship dwell times—the annual number of container cranes at ports. The noisy relationship supports the validity of our identification strategy, suggesting that our instrument is unlikely to be systematically related to endogenous, time-varying port investments aimed at particular ships.

Figure A.5. Residualized Plot of Container Cranes and Port Trade Exposure Instrument



***Notes*:** This figure plots the binned scatter plot and correlation between the residualized annual number of container cranes and port trade exposure instrument. The binned scatter plot is at the port-year level with 708 observations. Robust standard errors are clustered at the port-level.

Table A.7 presents robustness checks on our baseline results in Table 2.

Table. A.7. Congestion Elasticity of Port Traffic with respect to Ship Dwell Times: Robustness Checks

	(1) OLS	(2) OLS	(3) IV	(4) IV	(5) IV
Port Traffic	0.229		1.111	0.208	0.272
	(0.023)		(0.511)	(0.114)	(0.163)
Port Traffic × Before Mar 2020		0.091			
		(0.011)			
Port Traffic × After Mar 2020		0.120			
		(0.012)			
Day-Month-Year FE	✓	✓	✓	✓	✓
Port-Year FE	✓	✓	✓	✓	✓
Ship-Port FE	✓	✓	✓	✓	✓
March 2020 Period				Before	After
West Coast Ports	✓		✓		
Observations	22367	90516	22367	69917	20084
First Stage KP-F			10.01	143.42	176.34

***Notes*:** Robust standard errors in parentheses are two-way clustered at the ship and port level. All variables are in logs. Port traffic is the 28-day moving average of total daily net tonnage at the port. Weighted by ship net tonnage. Column (1) produces results using the OLS regression from Equation (4.3) when restricted to only west coast ports while Column (2) produces results using the OLS regression from Equation (4.3) with indicator variables for pre- and post-March 2020 to highlight the pandemic period. Column (3) produces the 2SLS regression using the IV from Equation (4.4) when restricted to only west coast ports. Column (4) produces the 2SLS regression using the IV from Equation (4.4) when restricted to the period before March 2020 while Column (5) is restricted to the period after March 2020. The estimates in Columns (4) and (5) are statistically significant at the 90% confidence level.

Table A.8 presents the OLS regression from Equation (4.3) using shorter periods of moving averages for port traffic. With shorter periods of moving averages, the ship dwell times is still positively correlated with the average tonnage at the port but the magnitudes of the coefficients are smaller.

Table. A.8. Ship Dwell Times and Port Traffic by Time Aggregations

	(1)	(2)	(3)	(4)
Port Traffic	0.0995	0.0800	0.0540	0.0246
	(0.0104)	(0.00850)	(0.0151)	(0.00692)
Day-Month-Year FE	✓	✓	✓	✓
Port-Year FE	✓	✓	✓	✓
Ship-PortFE	✓	✓	✓	✓
Moving Average (Days)	28	21	14	7
Observations	90516	90516	90515	90492
R^2	0.77	0.77	0.77	0.77
F	92.19	88.54	12.82	12.68

***Notes*:** Robust standard errors in parentheses are two-way clustered by ship and port. All variables are in logs. Column (1) estimates the elasticity using the 28-day moving average of total daily net tonnage at the port and is replicated from the baseline results in Column (2) Table 2. Column (2) presents the 21-day moving average, Column (3) presents the 14-day moving average, and Column (4) presents the 7-day moving average. Weighted by ship net tonnage.

B.5 Rail Intermodal Terminal Congestion and Port Traffic

In this subsection, we study how port congestion can impact the multimodal network. In particular, we focus on how port traffic affects the amount of time a rail car spends at the rail station that is local to that port. We estimate the following regression (Columns (3) and (4), Table A.9):

$$\ln \text{Rail Dwell Time}_{rpwy} = \beta_2 \ln \text{Port Traffic}_{pwy} + \gamma_{wy} + \phi_{rp} + \epsilon_{rpwy} \tag{B.4}$$

where Rail Dwell Time$_{rpwy}$ is the average number of hours a rail car spends at a rail station r that is in the vicinity of port p during week w in year y, Port Traffic$_{pwy}$ is the average amount of port traffic at port p for that same week w in year y,[3] Port traffic is measured in both net tons of ships as in our congestion elasticity approach, as well as the number of ships. γ_{wy} is week-year fixed effects, and ϕ_{rp} is rail station-port fixed effects. The key parameter of interest, β_2, captures the link between rail dwell times and its nearest port traffic. Standard errors are clustered at the port level.

The week-year fixed effects control for aggregate events that affects all rail stations. The rail-port fixed effects control for fixed characteristics at the rail-port level. These include time-invariant comparative advantage differences across ports that result in larger capacity trains servicing the rail stations close to these ports which mechanically take longer time to unload. It also includes fixed rail station characteristics and fixed port characteristics that take into account their geography. In order to match the local rail stations in our rail dwell times dataset to ports, we extend the port areas in order to capture nearby rail stations. The buffer area we used in our baseline result is 150km which captures 6 ports and 11 rail stations. This small set of stations and ports is due to data restrictions from the rail dwell times dataset. The rail companies are only required to report the dwell times for the top 10 largest of their rail stations, and these rail stations sometimes overlap in geographic location.

We find that a one percent increase in port traffic, measured in average net tonnage, is correlated with a statistically significant increase in rail dwell times by 0.05 percent (Column (3), Table A.9). This elasticity is robust to specifications with rail station fixed effects and port fixed effects separately (Column (1), Table A.9). This elasticity is also robust to an alternative measures of port traffic by using the average number of ships (Columns (4) and (2), Table A.9).

As a robustness check, we extend the buffer area around the ports to 200km which captures 9 ports and 16 rail stations. We find that our estimate has the same sign and is within one standard error of our baseline estimate (Column (5), Table A.9). However, the magnitude of this estimate is smaller and noisy, due to the impact of port traffic being more muted on rail stations that are further away. Subsequent increases to the buffer area correspondingly result in even smaller estimates. Due to the small sample size, we are unable causally identify the multimodal impact of port traffic on rail dwell times, but we are able to show that there are statistically significant and positive correlations between the two.

[3]This measure, as mentioned from the previous subsection, is at the daily level. In order to match the rail dwell time dataset, we aggregate it up to the weekly level. We start our week on a Monday since we observe in our data that most ships tend to enter a port on Mondays.

Table. A.9. Link between Rail Dwell Times and Nearest Port Traffic

	(1)	(2)	(3)	(4)	(5)
Nearest Port Traffic (Net Tons)	0.0457		0.0457		0.0282
	(0.0152)		(0.0152)		(0.0196)
Nearest Port Traffic (Ships)		0.0372		0.0372	
		(0.0143)		(0.0143)	
Port Buffer Area	150km	150km	150km	150km	200km
Week-Month-Year FE	✓	✓	✓	✓	✓
Rail Station-Port FE					✓
Rail Station FE	✓	✓			
Observations	3327	3327	3327	3327	4316
R^2	0.80	0.80	0.80	0.80	0.79
F	9.08	6.80	9.08	6.80	2.06

Notes: Robust standard errors in parentheses are two way clustered by port and rail station. All variables are in logs. Local rail stations are determined by a 150km or 200km buffer area around the ports as described in Appendix Section B.1.6.

B.6 Model Fit: Additional Results

This section presents further results that evaluate our model's ability in representing actual observed trade flows in the data across various transport modes. Table A.10 reports the univariate regressions between residualized observed trade flows and residualized model-predicted trade flows at both the aggregate and mode-specific levels (Columns (1) and (2) respectively). Columns (3) and (4) in Table A.10 demonstrates the gravity model implications our model by comparing the negative relationship between residualized observed trade and distance versus residualized model-predicted trade and distance.

Table. A.10. Model Fit

	(1)	(2)	(3)	(4)
	Residualized Observed Trade	Residualized Observed Trade	Residualized Observed Trade	Residualized Predicted Trade
Residualized Predicted Trade	1.22***			
	(0.03)			
Residualized Predicted Trade for Truck		1.20***		
		(0.03)		
Residualized Predicted Trade for Rail		1.34***		
		(0.07)		
Residualized Predicted Trade for Barges		1.02***		
		(0.26)		
Residualized Route Distance			-1.15***	-0.61***
			(0.07)	(0.02)
Conditional on origin and destination FE	✓	✓	✓	✓
Observations	14514	14514	14467	14467
R^2	0.62	0.62	0.60	0.41
F	1524.65	521.92	314.11	671.98

***Notes*:** This table compares the observed bilateral origin to destination mode-specific trade flows with the mode-specific trade flows predicted by the multimodal economic geography model based on observed traffic data along the transport network. Both the observed and predicted trade flow measures are in logs and residualized using origin and destination fixed effects, allowing for the comparison to come from similarities at the origin-destination pair-level. Column (1) compares the aggregated observed and predicted trade flows while Column (2) compares the mode-specific observed and predicted trade flows for all three transport modes (truck, rail, and barges). Columns (3) and (4) examine the gravity model implications of the model by comparing the relationship between the observed trade flows and distance (Column (3)) to the relationship between the predicted trade flows and distance (Column (4)). The route distance measure is also in logs and residualized using origin and destination fixed effects. Robust standard errors in parentheses are clustered by origin and destination cities. Weighted by trade weight in tons.

B.7 Gains from Improving Intermodal Terminals Counterfactuals: Additional Results

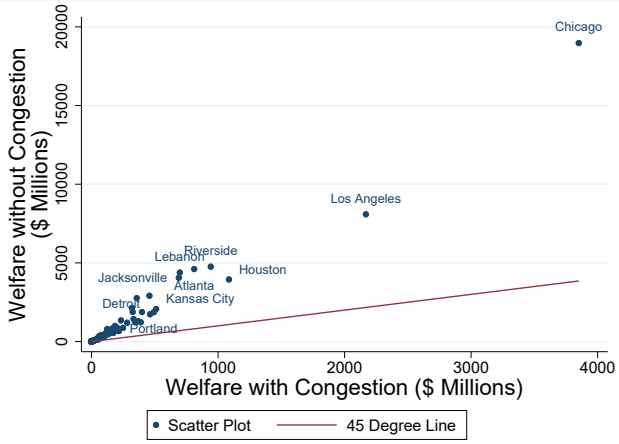
Table A.11 presents the top 30 list of intermodal terminals with the highest welfare impacts from lowering the transshipment cost in each terminal by 1 percent. Figure A.6 is a scatter plot comparing the welfare impacts from a 1 percent transshipment cost decrease at each terminal with and without congestion effects.

Table. A.11. Top 30 Ranking: Welfare Benefits of Improving Intermodal Terminals

	(1) CBSA Name	(2) Population	(3) Terminals	(4) Throughput	(5) ROI	(6) Benefit ($m)	(7) Cost ($m)
1	Chicago-Joliet-Naperville, IL-IN-WI	9368268	88	10368684	3.730	3851	814
2	Los Angeles-Long Beach-Santa Ana, CA	9639715	38	6836640	3.039	2168	537
3	Houston-Sugar Land-Baytown, TX	3133212	27	630300	24.782	1086	42
4	Riverside-San Bernardino-Ontario, CA	2173638	14	761760	10.999	942	79
5	Atlanta-Sandy Springs-Marietta, GA	1627623	28	1830840	4.643	811	144
6	Lebanon, PA	655561	1	793920	10.194	698	62
7	Jacksonville, FL	936317	18	797880	17.252	689	38
8	Kansas City, MO-KS	1767872	55	1088760	5.335	510	81
9	Portland-Vancouver-Hillsboro, OR-WA	1641801	30	424296	16.349	492	28
10	Detroit-Warren-Livonia, MI	2732964	31	557760	11.799	463	36
11	Memphis, TN-MS-AR	997862	28	1416120	3.722	457	97
12	Charlotte-Gastonia-Rock Hill, NC-SC	1502267	20	352440	13.404	399	28
13	Denver-Aurora-Broomfield, CO	2252276	19	248964	18.907	389	20
14	New York-Northern New Jersey-Long Island, NY-NJ-PA	14745610	29	1493556	2.160	371	117
15	Savannah, GA	250934	9	381840	10.908	357	30
16	Dallas-Fort Worth-Arlington, TX	4513776	13	1692480	1.608	347	133
17	Cincinnati-Middletown, OH-KY-IN	1744673	23	257160	15.447	332	20
18	Shreveport-Bossier City, LA	282413	4	390120	11.469	325	26
19	Birmingham-Hoover, AL	702669	23	395880	9.277	319	31
20	Allentown-Bethlehem-Easton, PA-NJ	891373	2	161760	21.166	282	13
21	St. Louis, MO-IL	1901086	38	592320	5.287	249	40
22	Charleston-North Charleston-Summerville, SC	480101	9	143400	26.965	233	8
23	Buffalo-Niagara Falls, NY	1150833	17	126600	15.004	221	14
24	San Francisco-Oakland-Fremont, CA	3863536	14	312936	1.233	215	96
25	Cleveland-Elyria-Mentor, OH	2175988	23	430440	5.305	213	34
26	Baltimore-Towson, MD	684543	16	145776	16.236	197	11
27	Syracuse, NY	664647	2	107040	29.231	196	6
28	Louisville/Jefferson County, KY-IN	1308587	20	185760	11.877	188	15
29	Laredo, TX	265111	3	130920	17.008	185	10
30	Stockton, CA	901910	11	864960	1.574	175	68

Notes: The table shows the thirty terminals where a one percent reduction of the switching cost generates the highest benefit. Column (1) indicates the core based statistical areas (CBSA) name of the node, which includes both metropolitan and micropolitan areas. The terminal's population, number of terminals, and rail throughput in TEUs are reported in Columns (2), (3), and (4) respectively. Column (5) shows the imputed return on investment (ROI), Column (6) calculates how much 2012 U.S. GDP would need to increase in order to match the overall welfare gain, while Column (7) presents the required cost of making this one percent cost decrease.

Figure A.6. Welfare Benefits with and without Congestion

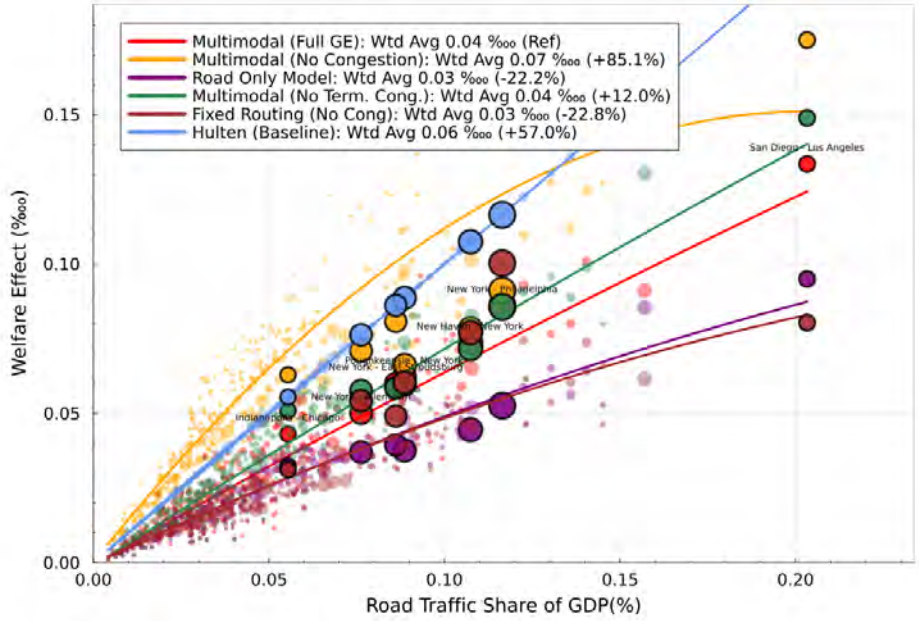


Notes: This figure presents a comparison of the welfare impacts from a 1 percent transshipment cost decrease at each terminal with and without congestion effects. See Figure 9 for the same comparison for the top 20 most impactful terminals.

B.8 Channel Decomposition Using Highway System Improvements: Additional Results

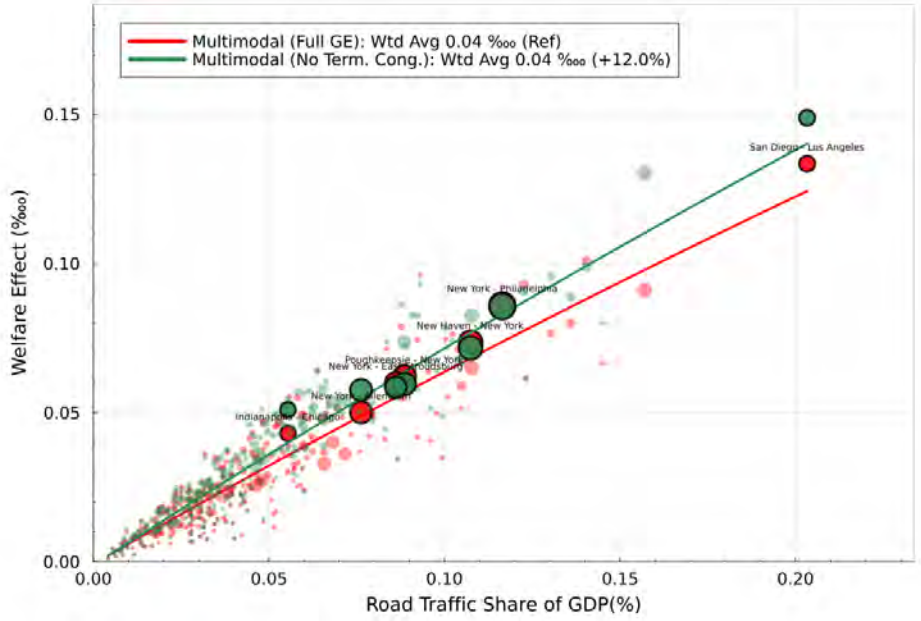
Figure A.7 shows the welfare effects of a one percent reduction in transport costs on each U.S. highway link and decomposes the underlying mechanisms by comparing the full multimodal model to counterfactual specifications that sequentially shut down road and terminal congestion (yellow), modal substitution (purple), terminal congestion only (green), endogenous routing (brown), and externalities (blue).

Figure A.7. Joint Decomposition of Main Channels: Welfare Effects of Highway Improvements



Notes: This figure reports the welfare effects of a 1% reduction in transport costs on each link of the U.S. highway segment and decomposes the main mechanisms in the model. The red line represents the full multimodal model, which includes road and terminal congestion, productivity and amenity spillovers (Hulten effects), multiple transport modes, and endogenous routing. Each of the other 5 lines shuts down one channel at a time: road and terminal congestion (yellow), modal substitution (unimodal road-only network, purple), terminal congestion only (green), endogenous routing (fixed routing, brown), and externalities (blue). The legend in each panel reports the weighted average welfare effect for each specification, with weights given by the GDP of origin and destination nodes of each link). The percentage in parentheses indicates the difference relative to the full multimodal model. For example, abstracting from congestion increases the average estimated welfare gain by 85.1% relative to the full model.

Figure A.8. Decomposition of Main Channels: Welfare Effects of Highway Improvements with and without Terminal Congestion



Notes: This figure reports the welfare effects of a 1% reduction in transport costs on each link of the U.S. highway segment and decomposes the main mechanisms in the model. The highway system consists of 704 links in total, each corresponding to a separate counterfactual. Each model specification evaluates the same set of 704 link-level counterfactuals. The red line represents the full multimodal model, which includes road and terminal congestion, productivity and amenity spillovers (Hulten effects), multiple transport modes, and endogenous routing. The green line represents a specification that shuts down terminal congestion. The legend in each panel reports the weighted average welfare effect for each specification, with weights given by the GDP of origin and destination nodes of each link). The percentage in parentheses indicates the difference relative to the full multimodal model: abstracting from terminal congestion increases the average estimated welfare gain by 12% relative to the full model. Figure A.7 presents all decomposition results jointly in a single figure.