Estimating Commuter Benefits of a New Transit System: Evidence from New York City's Ferry Service

Justin Tyndall

jtyndall@hawaii.edu
University of Hawai'i Economic Research Organization
and University of Hawai'i Department of Economics
2424 Maile Way, Saunders Hall 540, Honolulu, HI, USA, 96822

June 10, 2025

Abstract

This paper estimates the impact of a new transit system on worker outcomes, accounting for endogenous worker decisions. I examine the phased opening of New York City's commuter ferry system. I find evidence of a small but significant shift in commuting flows, towards routes with ferry service, driven by middle-to-high-income workers. I then propose and estimate a novel structural neighborhood choice model that recovers workers' valuation of ferry service and the aggregate effects of the system on employment. Higher-income workers display a stronger preference for the ferry. Ferry routes also match the location preferences of higher-income workers, allowing these workers to capture almost all direct benefits from the new system. Differing home and work location preferences across income groups largely determine who benefits from a new transit system.

Keywords: Transportation; Transit; Neighborhood choice; Ferry; Structural estimation

JEL: J64; R13; R23; R40; R58

1 Introduction

City planners face complex decisions when investing in transportation infrastructure. These choices shape urban form, influence long-term economic growth, and carry significant distributional implications. The selection of a transit mode affects the overall efficiency of the system, while the location of service determines who benefits. Adding to this complexity, workers may relocate or adjust their commuting behavior in response to changes in transit options. This paper develops an empirical framework for estimating the impacts of new transit infrastructure in settings where workers endogenously choose both where to live and where to work. I apply this framework to analyze the recent expansion of New York City's ferry system, using detailed commuter flow data to evaluate labor market effects.

Beginning in 2011, New York City undertook a major expansion of its ferry network to enhance commuter mobility. This setting provides a valuable opportunity to examine how new transit links affect commuting and residential choices. Unlike subways and buses, which have been widely studied, ferries offer a relatively understudied mode of transit that can directly connect otherwise isolated neighborhoods. I begin by estimating the causal effect of new ferry connections on commuting flows. Ferry service increased commuting between connected neighborhoods by an estimated 2,600 workers, driven by higher-income, middle-aged commuters. The total number of workers living near ferry terminals grew, particularly for workers in the FIRE (Finance, Insurance, and Real Estate) and business-service sectors. I then use a quantitative spatial model to value ferry access and disentangle sorting from labor supply effects. The model attributes 700 new city-wide jobs to ferry service and 1,900 additional connections to sorting, yielding a 0.02% increase in the city-wide employment rate.

I innovate on existing structural neighborhood choice methodology by introducing a simple model that accounts for location preferences across worker types and solves for an equilibrium that is consistent with observed commuter flow changes. I argue the proposed model can be simply estimated and requires limited data inputs. The proposed methodology could be applied in similar settings to estimate transit benefits.

The analysis yields some new and novel findings. First, worker location preferences are important in determining the distribution of system benefits, despite the ability of workers to relocate. I find that ferry uptake is heavily concentrated among middle-to-high-income workers who already had a preference for routes served by the ferry and who display a relative preference for ferry commuting. Second, I estimate the aggregate labor market impact of the ferry, accounting for equilibrium rent changes. I find a marginal impact on city-wide employment with essentially no impact on aggregate employment of low-income workers.

This paper fits closely into a growing literature of quantitative spatial equilibrium models. The recent structural modeling approaches are empirical applications of the basic urban spatial model (Alonso, 1964; Muth, 1969; Mills, 1967; Fujita and Ogawa, 1982), and urban choice model (Tiebout, 1956). Anas (1981) and Epple and Sieg (1999) extended the discrete choice framework of McFadden (1973), modeling urban neighborhood choice in a discrete choice framework. Bayer et al. (2004) provided further methodological extensions and Bayer and McMillan (2012) explicitly reconciled neighborhood choice modeling within a Tiebout (1956) framework. Other important applications of neighborhood choice modeling include Sieg et al. (2004); Bayer et al. (2007); Ferreyra (2007), and Ahlfeldt et al. (2015).

Some recent, closely related work has specifically applied quantitative spatial equilibrium models to estimating the benefits of new transit systems. Severen (2023) estimated welfare effects of the Los Angeles subway system using a discrete choice model. The author develops a method of using route-level fixed effects applied to panel commute flow data. Tyndall (2021) analyzed the effects of US light rail transit systems on commuter flows and neighborhood change, combining a discrete choice framework with parameter estimates generated from an instrumental variable regression analysis. I will make use of reduced-form regression results to parameterize a model and follow some modeling assumptions made in Tyndall (2021). Chernoff and Craig (2022) provided an application for a rail transit expansion in Vancouver, Canada, examining distributional effects across worker types. Mo (2023) modeled household responses to a changing road network in Xiamen, China.

I contribute a further extension to urban discrete choice modeling. I develop a quasi-difference-in-difference setup based on route-level fixed effects. The fixed effects approach constricts the model to closely adhere to observed commute flow data and eliminates the need to directly estimate a bilateral commuting time matrix. Dingel and Tintelnot (2020) provided a methodological review of fixed-effects-based structural estimation strategies, arguing granular datasets with a large number of fixed effects and sparse data coverage can lead to overfitted models with biased results. The approach of this paper, to pool data across 18 years and estimate time-invariant route fixed effects, aims to overcome this issue. Overall, I attempt to synthesize past structural estimation approaches to produce a model that is simple and tractable with limited data requirements.

This paper is particularly concerned with the impact of transit on establishing labor market connections. Kain (1968) introduced the concept of spatial mismatch, arguing that localized unemployment could be driven by insufficient spatial access to job opportunities. Andersson et al. (2018) provided contemporary evidence of spatial mismatch in the US. Some studies have attempted to establish a causal relationship between job access and employment through natural experiments. Holzer et al. (2003) looked at reverse commuters using a new transit line in the San Francisco area. Tyndall (2017) used New York City subway system closures due to Hurricane Sandy as a source of random variation in transit access. Both studies found a positive, causal link between transit access and employment.

My setting concerns a commuter ferry system. While buses or trains connect neighborhoods that are arranged linearly, ferry routes may connect neighborhoods that were otherwise spatially isolated from one another. The connecting of formerly disparate neighborhoods through transit provides a cleaner environment to estimate commuter impacts than would be possible for a bus or rail route. Spatially separated neighborhoods are less likely to share underlying, unobserved characteristics or trends, which may contaminate difference-in-difference style designs. Prior literature has responded to spatial endogeneity concerns through instrumental variables (see for example, Chernoff and Craig (2022); Severen (2023); Tyndall (2021)).

Across the US, there are 44 operating commuter ferry services, which provide 90 million passenger trips per year.[1] There is limited economic literature examining the effects of ferry service. Sandell (2017) examined a reconfiguration of Sydney, Australia's ferry system, and argued that route selection that accounts for route-level, rather than neighborhood-level, travel demand is important to creating an efficient system. Thompson et al. (2006) discussed the urban form consequences of ferry service in New York City, and argued that ferries may play an important role in waterfront Transit Oriented Development (TOD) projects in coastal US cities. Schreurs et al. (2023) provided a detailed description of New York City's ferry expansion and argued that ferry terminal locations were selected based on local real estate development opportunities and that the terminals contributed to local gentrification.

The paper will proceed as follows. Section 2 provides a background of the New York City commuter ferry service. Section 3 describes the data used in the paper. Section

[1] American Public Transportation Association 2021 Factbook.

4 presents reduced form estimates of the ferry's impact on bilateral commuter flows. Section 5 presents a structural approach to estimate underlying preference parameters for ferry service and aggregate employment effects of the system and Section 6 concludes.

2 Ferry Service in New York City

Commuter ferry service in New York City has a long history. Before the construction of a bridge and tunnel network connecting Manhattan to Long Island and New Jersey, privately operated ferry services were a vital link in the region. As bridges and tunnels were completed in the late 1800s and early 1900s ferry services were generally phased out.[2] Vilain et al. (2012) provides a useful discussion of the history of ferry services in New York City up to the 2011 opening of the East River Ferry, which is the first expansion route considered in this paper.

This paper will analyze the recent revival of commuter ferry service in New York City, a system now known as NYC Ferry. Figure 1 provides a public map of the six ferry routes operated as of May 2019.[3] The opening dates of the six new ferry routes were staggered from 2011-2018. Figure 2 provides a timeline of route opening dates. The East River route opened in 2011, significantly earlier than the other five routes, which began operating in either 2017 or 2018.[4] I make use of the staggered opening dates as a source of variation in my empirical identification approach.

The New York City ferry system has been the topic of sustained political and public attention. In 2011, under the Bloomberg mayoral administration, the city began operating the East River Ferry route. A comprehensive report on possible ferry system expansion was released in 2011 (NYCEDC, 2011) and updated in 2013 (NYCEDC, 2013). Subsequently, the broader expansion of ferry service became a cornerstone of transit investment efforts under the de Blasio mayoral administration. In 2015 it was announced that the ferry service would be expanded to include an additional five routes with the system being re-branded as NYC Ferry. Investment in the ferry system was justified as a way to relieve stress on the overburdened and aging subway system. An explicit goal of the system was to expand employment opportunities for disadvantaged

[2] An exception is the Staten Island Ferry, which connects the north of Staten Island with Lower Manhattan and has provided commuter ferry service continuously since 1817. I do not consider the Staten Island Ferry in the analysis, as its impact is constant across the study period.

[3] I ignore the ferry route to Governors Island, which is not a commuter ferry but provides recreational access to the park on Governors Island.

[4] The Lower East Side route closed in 2020. In 2021, a route serving Staten Island and the west side of Manhattan opened. Both of these events occur after my study period.

Figure 1: Map of Ferry Service

A public map was disseminated by the ferry operator in 2019, showing the routes in operation.

and isolated workers. For example, at a press conference announcing system expansion, Mayor de Blasio stated; "*If you can't get to a job interview or a job...you just don't have as much opportunity to get ahead economically. We don't want to see that happen.*"[5]

Many local media outlets and transit advocates responded to plans by arguing that scarce public transit funds could be more effectively invested in the existing subway and

[5]Quoted from a July 26, 2017 Press Conference. NYC Mayor's Office. *Mayor de Blasio Makes Announcement About NYC Ferry Service.*

Figure 2: Opening Dates of Ferry Lines

2011	East River.
2017	Astoria. Rockaway. South Brooklyn.
2018	Lower East Side. Soundview.

A timeline of ferry route openings is shown for the six routes operating during the study period.

bus systems.[6] Given the locations of planned routes, concerns were raised that ferry service would primarily serve high-income residents of the city and fail to meaningfully improve transit for those most reliant. For most routes covered by the NYC Ferry system, the commute time is shorter via subway than by ferry. While NYC Ferry was uncompetitive with the subway system in terms of estimated trip duration, the ferry system was less prone to delays than other modes and provided a more comfortable experience for riders.

The most important ferry node is the Wall Street terminal, which connects to all six ferry routes. The area surrounding the Wall Street station has a high concentration of jobs, particularly high-income jobs. In 2010, 71% of jobs in the Wall Street neighborhood paid more than $40,000 annually, while the rate in the rest of the city was 50%.[7] Including the Wall Street station, the system encompasses five stations on Manhattan Island, nine stations in Brooklyn, four stations in Queens, one station in the Bronx, and one station on Roosevelt Island.

Ferry hours and headways vary by route, but ferries generally operate from 7 a.m. to 9 p.m. and run with 20-minute headways on most routes. Ferry fares were set at $2.75, matching subway fares.[8] However, the ferry system utilized a separate payment system and therefore did not allow for free or discounted transfers between the ferry

[6]For an example of media coverage see: *A Ferry Subsidy of $24.75 a Ride? New York City's Costs Are Ballooning.* New York Times. April 17, 2019.

[7]I use census tract boundaries to approximate the Wall Street area, which I consider to be all of Manhattan south of Chambers Street. Wage data is taken from the 2010 LEHD LODES Worker Area Characteristics file.

[8]In 2022, standard ferry fares were increased to $4.00, but this change occurred after the study period of this paper.

and other modes of public transit. While fares rarely cover the operating costs of public transit systems, the NYC Ferry system required particularly high public subsidies. A nonpartisan audit of ferry operations in 2018 found that every passenger trip required a $10.73 government subsidy. For comparison, subway trips required a $1.05 per ride subsidy, and commuter rail services required a $6 per ride subsidy (Campion, 2019).

In 2018, the annual operating cost of NYC Ferry was reported to be $56.7 million. Capital expenditures to establish the system were estimated to total $639 million. Debt servicing on the bonds needed to fund the capital expenditures was estimated to cost the city $48.6 million per year, for 20 years (Campion, 2019). By adding annual operating expenses to debt servicing I consider the annual public cost of the ferry system to be $105 million.

Despite a significant revival in ferry service, the share of commuters in New York City who commute by ferry is small. According to data from 2016-2020, only 0.3% of New York City commuters used a ferry as their primary mode of commuting.[9] Table 1 provides commuter mode shares. The subway is the most popular mode of commuting in the city, and is the primary mode for 41% of commuters, meaning there are roughly 140 subway commuters for every ferry commuter. Ferry ridership in New York City has experienced growth in recent years. From data spanning 2005-2009 only 0.23% of commuters used a ferry as their primary mode, while the 2016-2020 data shows 0.33%, marking 50% growth. Figure 3 graphs the growth in ferry commuter mode share.

Actual ridership data from NYC Ferry showed relatively strong numbers, with some routes hitting ridership capacity at peak hours. Ferry boardings across the system on a weekday were roughly 14,000 in 2019 (NYC Ferry, 2019). On-board surveys of ferry users have found nearly 60% of trips are for purposes other than commuting (New York City Economic Development Corporation, 2019).

The recent expansion of New York City's ferry service is consequential in terms of the system's costs as well as the public and political attention the system garnered. Evaluating the labor market consequences of the New York City ferry service is important to understand the impact of this large public investment.

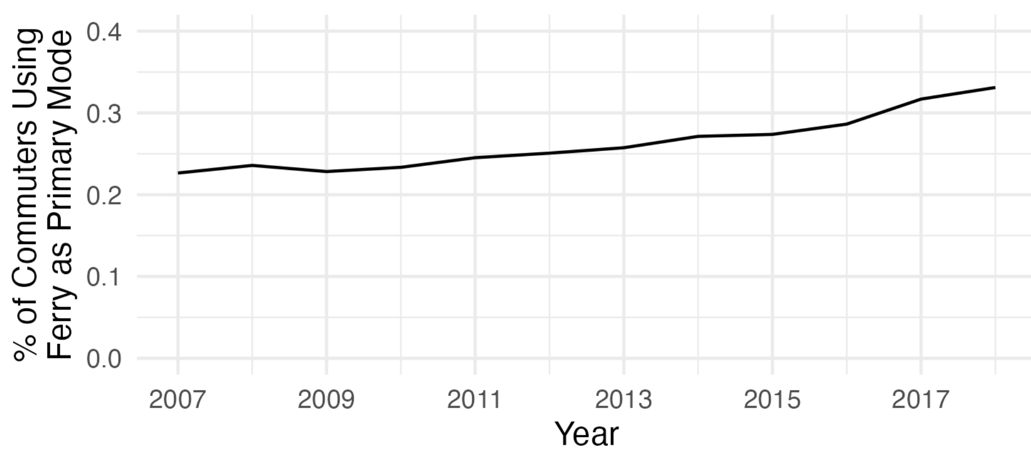
[9] The ACS question asks about a commuter's "primary mode" of transport. If the commuter uses multiple modes to complete their commute, they are asked to report the mode "used for most of the distance." Therefore, the measure undercounts the number of people who use the ferry as one component of a long commute.

Table 1: New York City Commuter Mode Shares

	Share of Commuters
Private Vehicle	26.77%
Drove Alone	22.31%
Carpooled	4.46%
Public Transportation	52.82%
Bus	9.76%
Subway	41.24%
Commuter Rail	1.34%
Ferryboat	**0.33%**
Taxicab	1.10%
Motorcycle	0.08%
Bicycle	1.32%
Walked	9.78%
Worked From Home	7.40%
Other Means	0.88%

Data from 2020 5-year American Community Survey.

Figure 3: Ferry Mode Share in New York City



Data is from the 5-year American Community Surveys. Data is plotted according to the center year of the survey. For example, data for the 2018 point includes survey responses from 2016-2020.

3 Data

I construct a unique data set from multiple sources. Using a full list of 2010 New York City census tracts, I generate a matrix of every possible home-location work-location pair. I then expand this set to include one route-level observation per year across the study period, which covers 2002-2019.[10] New York City contains 2,167 tracts, though a small number of these contain either no housing or no employment across the whole study period meaning they are omitted as possible commute routes. The final set is a balanced panel, containing 82,581,120 route-by-year observations, including 4,587,840 unique routes, 18 unique years, 2,124 unique home tracts, and 2,160 unique work tracts.

To incorporate data on commuter flows, I use the Longitudinal Employer-Household Dynamics, Origin-Destination Employment Statistics (LODES). The data is provided annually as a matrix of bilateral commuter flows at the census block level, recording the number of workers who complete commutes between any two particular blocks. I collapse the data to the census tract level, generating a tract-level matrix of commuter flows for each year from 2002-2019. I join the tract-level LODES matrices onto the full set of all 83 million tract-to-tract by year observations. I add zero values to routes that had no commuters reported in the LODES data. Some routes contain zero reported commuters throughout all years of the study period. I discuss the issue of a large number of zeros in the description of the empirical estimation strategy. To avoid expanding my data set beyond New York City, the analysis ignores any worker in the LODES data who either lives or works outside of the city boundaries.

LODES includes breakouts by worker type. For income, LODES can be grouped into workers earning above or below $40,000 annually. The wide income buckets are a limitation of the LODES data. Throughout the paper, I refer to low-income workers as those earning below $40,000 and I refer to those earning above $40,000 as either middle-to-high-income workers or higher-income workers. For individual income, $40,000 corresponds to roughly the 54th percentile of the income distribution in New York City during the study period.[11]

For every route-year observation, I determine whether that route-year is served by a ferry connection. I first geocode ferry terminal locations from a list provided by

[10] I choose to end the study period at 2019 to avoid the potentially confounding effects of the COVID-19 pandemic. From 2019 to 2020 unemployment in NYC roughly tripled, and ferry ridership plummeted.

[11] 2013 5-year American Community Survey, Integrated Public Use Microdata Series.

the City of New York. I generate circular buffers around ferry terminals to identify an area of pedestrian access. In the main specification, I use a 200-meter buffer. I overlay terminal buffers on the census tract shapefile and identify all census tracts that are overlapped by the local buffer of a particular terminal. If the tract and buffer overlap at all I consider the tract to be treated by that ferry terminal. I consult the route map (Figure 1) to identify every pair of terminals that are served by a common ferry route. I use the opening date of each of the six newly established ferry routes. The variable for an active ferry connection takes a value of one if the route is served by a common ferry route, meaning the tracts are both within the buffer of a terminal that accommodates the common route[12] and the ferry is open to the public in that year.[13]

Figure 4 provides a map of the census tracts that were connected by ferry service by the end of the study period. The majority of connected tracts are located along the East River in Brooklyn and Queens. According to tract-to-tract flow data from the Census Transportation Planning Products (2012-2016) the specific routes connected by ferry service have a public transit mode share of 48%, comparable to the city-wide rate (53%).

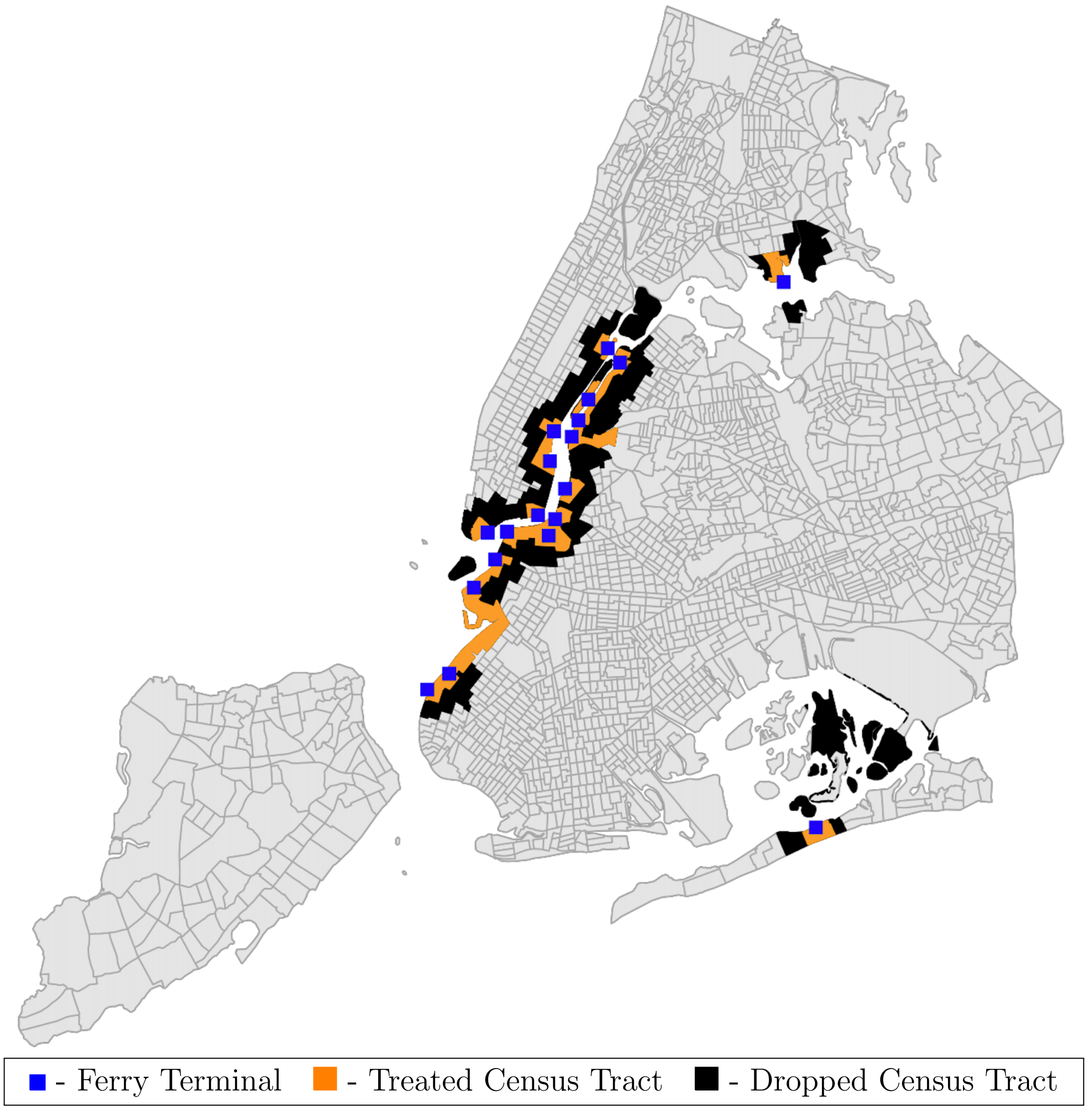
Census tracts that are not directly adjacent to a ferry terminal, but still relatively near a terminal, may be partially treated as some workers may choose to walk farther to use the ferry. When conducting analysis I drop nearby tracts to reduce the SUTVA issue of partially treated routes. In the main analysis, I drop routes that include tracts that are not intersected by the 200-meter buffers but are within a 1,000-meter buffer of a terminal. I drop these partially treated routes across the entire study period to maintain a balanced panel. These dropped routes represent only 0.3% of the original sample. I show these tracts in Figure 4.

Table 2 provides summary statistics. Across the final sample, the average route has 0.58 commuters, where an average of 0.32 were low-income and 0.25 were middle-to-high-income. Only a small fraction of routes are connected by the ferry service. Across the whole study period, 0.0035% of routes are connected, which accounts for 2,859 route-year observations. At the end of the study period (2019), I identify 502

[12] I ignore the possibility that commuters may transfer across multiple ferry routes. Because of relatively long headways, completing a commute using multiple ferry routes is unlikely to be a reasonable commute in most cases.

[13] Because ferry routes do not necessarily open at the start of a year, and the LODES data is aggregated annually, I consider tracts connected if the ferry connection opened either in a preceding year or at some point during that year. I choose to consider partially treated years as treated rather than untreated in part because anticipation effects among workers mean that labor market responses may precede the actual opening of the ferry connection.

Figure 4: Census Tracts Treated by a Ferry Route



Tracts shown in orange are connected by at least one ferry route. Treated tracts are those that overlap a 200-meter buffer centered on a ferry terminal. Dropped tracts are those that are not overlapped by a 200-meter buffer but are overlapped by a 1,000-meter buffer.

connected routes, which comprise 0.011% of that year's routes.

In secondary analysis, I perform census tract-level regressions to test the average effect of a ferry connection on neighborhood conditions. For tract-level analysis, I use the LEHD LODES Resident Area Characteristics and Worker Area Characteristics files. These data files provide more detailed information on commuters but are identified by

Table 2: Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.
Workers	0.577	3.234	0	1166
Low-income workers	0.324	1.677	0	823
Middle-to-high-income workers	0.253	2.051	0	624
Young workers	0.140	0.950	0	590
Middle-aged workers	0.328	1.929	0	571
Older workers	0.108	0.748	0	287
Ferry link dummy	0.000035	0.005894	0	1
N		82,294,092		

Each observation corresponds to a unique route-year combination. Routes are dropped if they include a dropped tract as identified in Figure 4. The data set is a balanced panel of 4,571,894 routes and 18 years.

home and work location respectively, rather than assigning workers to specific home-work pairs. Similar to the route level data, I drop tracts that are partially treated (Figure 4).

In the structural estimation section, I will make use of tract-level ACS data, particularly median rents. I take data from the mid-point of the study period by using the 2013 5-year ACS.

4 Estimating the Commuter Flow Impact

In this section, I estimate the growth in route-level commuter flows caused by a ferry connection. When two tracts become connected by the ferry, the commuting cost between those tracts is reduced and this may attract new commuters. Workers may shift towards the ferry-connected routes by changing their home location, their work location, or entering the labor force. Workers may substitute away from an existing or prospective job. I explore the mechanisms of neighborhood sorting, extensive-margin employment decisions, and aggregate employment effects more completely in the subsequent structural estimation section. Accurately estimating the partial effect of a ferry connection on the number of commuters completing that route provides a test for whether the ferry service meaningfully affected commute flows and will be an important parameter in the structural estimation methodology.

4.1 Regression Methodology

I estimate new labor market connections through a difference-in-difference setup. The main regression specification is shown in Equation 1. C_{rt} is the number of commuters who commute along route r in year t. F_{rt} is a dummy variable that takes a value of one if route r was connected by the ferry system in year t. Φ_r is a vector of route fixed effects and Ψ_t is a vector of year fixed effects.

$$C_{rt} = \beta_0 + \beta_1 F_{rt} + \Phi_r + \Psi_t + \varepsilon_{rt} \tag{1}$$

The inclusion of route fixed effects absorbs any average difference in commute popularity across routes. Time-invariant, tract-level variation is nested within the route-level fixed effects, so Φ_r also controls for location-specific variation. Year fixed effects absorb any city-wide changes in commuter flows over time. Identification of the effect of the ferry connection makes use of only the temporal changes in flows that correspond to the timing of ferry route openings. The coefficient of interest is β_1, which captures the partial effect of a ferry connection on the number of commuters using that route, relative to non-treated routes. In some specifications, rather than capture the total number of workers, C_{rt} will capture the number of workers of a specific type, for example, low-income workers.

I assume a parallel trend where the growth in commuting flows along treated and control routes would have proceeded similarly if not for the introduction of the ferry system. While some commuting infrastructure changed during the study period, for example, the Second Ave Subway extension[14] (Gupta et al., 2022) and the introduction of some express bus routes (Tyndall, 2018), I assume the impacts are spatially orthogonal to the impact of ferry routes. I include an unconditional parallel trends figure in the appendix (Figure A1).

Recent developments in difference-in-difference methodology have yielded improved estimators for cases with staggered treatment (Goodman-Bacon, 2021; Callaway and Sant'Anna, 2021), which is the case here. The issues raised in the literature apply particularly to the current scenario because the effect of treatment by a ferry connection

[14] I drop a set of tracts from analysis that are more than 200-meters but less than 1,000-meters from a ferry terminal (Figure 4). Coincidentally, the eight census tracts that are directly bisected by the Second Ave Subway expansion are all within the 200-meter to 1,000-meter range and are dropped from analysis. Therefore, the direct effect of the subway expansion on worker flows to or from tracts adjacent to the new subway stations will not affect the analysis. Spillover effects of the subway expansion to other tracts could affect results. I assume these effects are orthogonal to the effect of the ferry system.

may trigger a shift in commuter *growth* over time, rather than a discrete level change. As outlined in Goodman-Bacon (2021) the staggered timing of treatments would result in a biased estimate when using a conventional two-way fixed-effect (TWFE) estimator. I use the estimation strategy for staggered treatment in difference-in-difference settings outlined in Callaway and Sant'Anna (2021), which is not subject to the bias identified in Goodman-Bacon (2021).[15]

In addition to estimating the effect of a ferry connection on route-level commuter flows, I estimate the effect of being connected to the ferry network on a census tract's number of locally residing workers and the number of local jobs. Equation 2 provides the regression equation for this tract-level analysis.

$$Y_{jt} = \alpha_0 + \alpha_1 F_{jt} + \Phi_j + \Psi_t + \varepsilon_{jt} \tag{2}$$

Y_{jt} is the number of workers or jobs located in tract j in year t, depending on what is being tested. F_{jt} is a dummy variable that takes a value of one if that tract is treated by an active ferry terminal in that year, meaning there is an open ferry terminal in the tract or within 200-meters of its boundary. Φ_j is a vector of tract fixed effects, and Ψ_t is a vector of year fixed effects. The coefficient of interest is α_1, which captures the partial effect of a local ferry terminal on the number of locally residing workers or jobs. I estimate Equation 2 using the Callaway and Sant'Anna (2021) method.

I estimate clustered standard errors for both the Equation 1 and 2 specifications. Errors are clustered at the route level for Equation 1 and the tract level for Equation 2.

4.2 Regression Results

Table 3, column 1 provides the main regression results (Equation 1). I find that when two tracts were directly connected through the ferry system the number of workers commuting between those tracts increased significantly, by 2.4 workers, on average. Among 502 treated routes, the average commuter flow in the pretreatment period was 6.2, meaning the ferry connection increased commute flows along these routes by 39%. For comparison, Severen (2023) found that tracts connected by LA's metro system experienced a 10-22% increase in commuting. As noted, subway or bus lines connect

[15] As a robustness check, I provide alternative results using a standard two-way fixed effect (TWFE) approach, a TWFE model that includes location-specific linear time-trends, a TWFE model with location-by-year fixed effects, and results from a Poisson pseudo-maximum likelihood estimator.

areas that are linearly arranged and were therefore relatively accessible to one another to begin with. Ferry connections link neighborhoods that were previously more isolated from one another and may therefore represent a more significant reduction in commute cost between tract pairs. Beyond potential time savings, the ferry may be providing benefits in terms of reducing trip time uncertainty or improving rider comfort.

Table 3: Effect of Ferry Connection on Commuter Flow by Income Group and Specification

	(1) Main	(2) TWFE	(3) TWFE + Location Time Trends	(4) TWFE + Tract by Year FE	(5) Poisson PML	(6) Waterfront Only	(7) Ferry-Treated Only	(8) TWFE Spillovers
Panel A: All Workers								
Ferry link dummy	2.408**	3.006**	2.140**	2.052**	0.201**	2.543**	2.322**	2.041**
(0–200m)	(0.392)	(0.541)	(0.512)	(0.504)	(0.060)	(0.405)	(0.392)	(0.502)
Ferry link dummy								0.467**
(200–1000m)								(0.083)
Pre-treatment avg.	6.181	6.181	6.181	6.181	6.181	10.036	10.095	–
% change	+39.0%	+48.6%	+34.6%	+33.2%	+22.3%	+25.3%	+23.0%	–
Panel B: Low-Income Workers								
Ferry link dummy	0.243*	0.161	0.047	0.001	0.070	0.277*	0.241*	-0.005
(0–200m)	(0.096)	(0.095)	(0.092)	(0.088)	(0.055)	(0.100)	(0.096)	(0.089)
Ferry link dummy								-0.081
(200–1000m)								(0.052)
Pre-treatment avg.	1.656	1.656	1.656	1.656	1.656	3.434	3.397	–
% change	+14.7%	+9.7%	+2.8%	+0.0%	+7.3%	+8.1%	+7.1%	–
Panel C: Middle-to-High-Income Workers								
Ferry link dummy	2.165**	2.845**	2.094**	2.052**	0.069	2.266**	2.081**	2.047**
(0–200m)	(0.364)	(0.507)	(0.471)	(0.467)	(0.069)	(0.377)	(0.363)	(0.464)
Ferry link dummy								0.548**
(200–1000m)								(0.073)
Pre-treatment avg.	4.525	4.525	4.525	4.525	4.525	6.602	6.698	–
% change	+47.8%	+62.9%	+46.3%	+45.3%	+7.1%	+34.3%	+31.1%	–

Significance levels: $*$: 5% $**$: 1%. The results from 24 separate regressions are shown. The estimation approach of columns 1, 6 and 7 follow Callaway and Sant'Anna (2021), columns 2, 3, 4 and 8 use a standard two-way fixed effects (TWFE) set-up and column 5 uses a Poisson pseudo maximum likelihood regression. Standard errors are shown in parentheses. N=82,294,092 for columns 1-5 and 8; N=10,931,112 for column 6; N=1,149,048 for column 7.

Table 3 provides results separately for low-income (Panel B) and middle-to-high-income workers (Panel C). The workforce in New York City is roughly evenly split between the two income groups, with 56% of workers being in the low-income group. However, among commute routes that would gain a ferry connection, the pre-ferry flows were 73% middle-to-high-income workers, demonstrating that middle-to-high-income workers were disproportionately common along the routes selected for ferry service, even before service began. In addition to middle-to-high-income workers already having commute routes that were more likely to benefit from ferry service, I find more middle-to-high-income workers altered their commutes to take advantage of ferry service. A ferry connection increased the number of low-income workers on a route by 0.24, but increased the number of higher-income workers by 2.17. Therefore, middle-to-high-income workers represent 90% of induced commuter flows. The over-representation of higher-income commuters in pre-ferry flows and their over-representation in induced commuter flows are consistent with public concern that the ferry system would disproportionately serve higher-income groups.

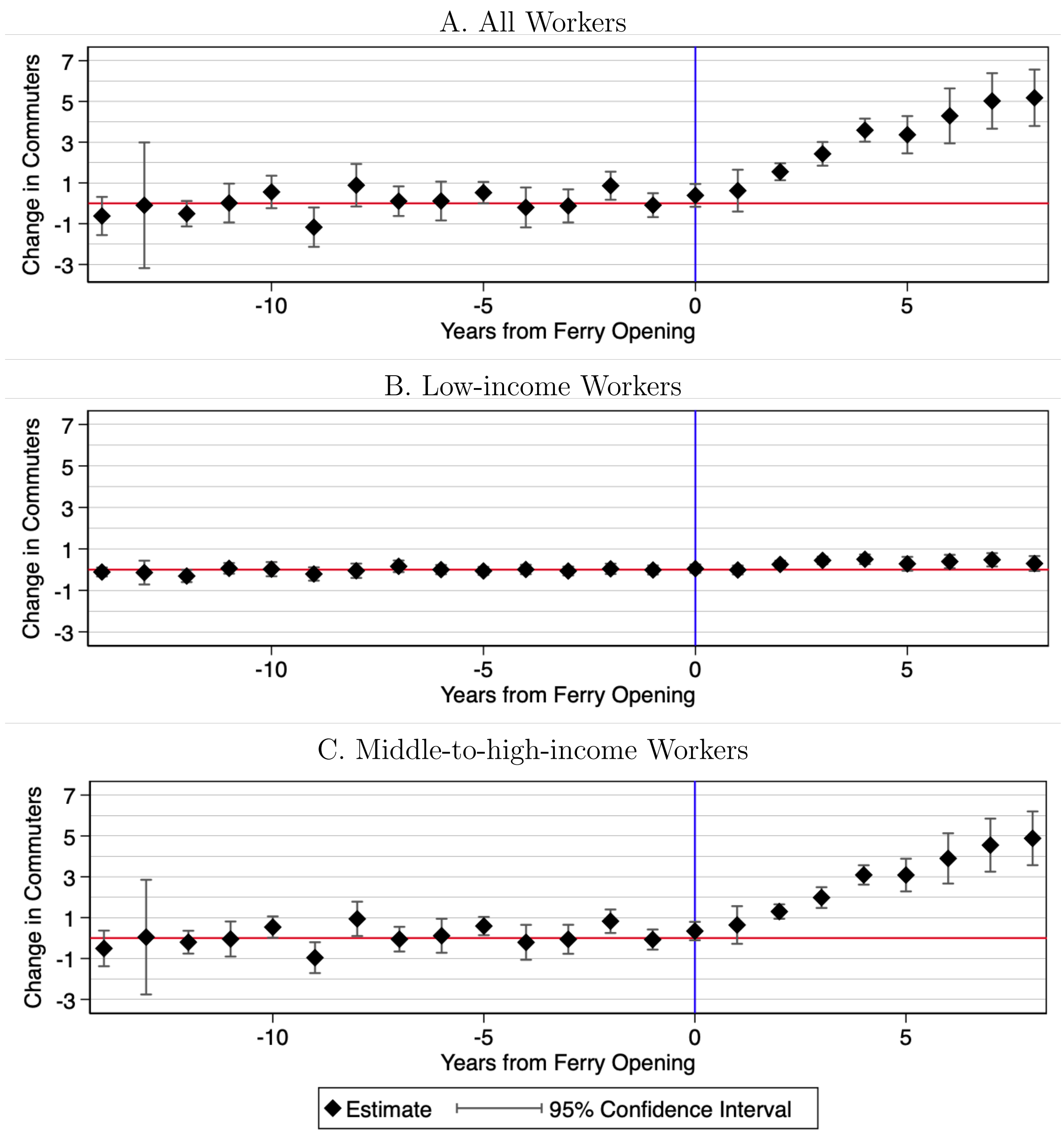
Figure 5 visualizes the main result as an event study, shifting the staggered treatment years so that 0 indicates the first year ferry service was available for a route. I again adopt the methods described in Callaway and Sant'Anna (2021) to construct the event study estimates. Panel A provides results for all workers, while Panels B and C provide results for low as well as middle-to-high-income workers respectively. Before treatment with ferry service, I find no evidence of a sustained difference in trends between the treated and control routes, which provides evidence for the validity of the parallel trend assumption. After treatment, I find a statistically significant effect beginning two years after the ferry opening. I find evidence of a change in the growth of commuters that is specific to ferry-treated routes, as opposed to an immediate level shift. The effect appears to stabilize in the final years observed, with an average increase of about five commuters per connected tract pair.

By 2019, the ferry system connected 502 tract pairs. The final point estimate in Figure 5A equals 5.18, which I take to be the long-run effect of a connection. The estimate suggests that routes with ferry service gained 2,600 workers, relative to routes untreated by a ferry connection. The impact of ferry service on the overall city labor force was small; 2,600 workers represent only 0.1% of New York City's total labor force. Additional commuters may have benefited from ferry service by switching modes, for example from subway to ferry, but not switching their origin-destination pair on account of ferry service.

Figure 5 Panels B and C show that the effect of a ferry connection on worker flows is almost entirely among middle-to-high-income workers. While I estimate a significantly positive effect for low-income workers in some years, the magnitude is negligible. By the final period of the event study, the effect on low-income flows is not statistically significant.

The estimate is affected by a possible SUTVA violation. If the workers moving to the ferry-connected route were drawn from a control route, they contribute to the effect by both increasing the treatment route flows and decreasing the control route

Figure 5: Event Study of Ferry Treatment Effect on Route Level Flows



Estimates and confidence intervals are calculated using the methods described in Callaway and Sant'Anna (2021).

flows. This effect would bias the estimates upwards. The β_1 estimate is effectively an upper bound on the job creation impact of the ferry system. While some of these new flows could be from previously unemployed workers, a portion may be workers who would otherwise have been employed along a different commuting route. I account for the effect of flows shifting between the treated and control units in the subsequent structural model.

As noted in the prior section, I estimate the model using the methods outlined in Callaway and Sant'Anna (2021) for staggered treatment difference-in-difference estimation. Table 3 provides numerous alternative specifications to establish the robustness of the main result. Column 2 repeats the main regression analysis (Equation 1) but estimates the equation using a traditional TWFE estimator rather than the Callaway and Sant'Anna (2021) approach. Column 3 again uses a TWFE model but also includes home and work tract linear time trend controls. Column 4 provides a more flexible TWFE estimator by including home tract by year and work tract by year fixed effects. All TWFE results are highly consistent with the main (Column 1) estimates. The choice to use a linear model, despite the outcome variable containing zeros and having significant skew, may introduce a separate source of bias. Column 5 estimates the main model again but uses a Poisson pseudo maximum likelihood estimator. The Poisson result for the overall change in commuter flow (Panel A) is consistent with the main result. However, the income-group specific estimates are no longer statistically significant.

As a control group, I draw from all route-by-year observations that were not treated by a ferry. While using the full sample makes use of all statistical variation available, the use of such a broad control group may include observations that are very different from the set of waterfront tract pairs that are treated. Plausibly, the ferry-treated routes could be subject to a different set of non-ferry-related shocks over the study period, which could bias estimation. As a robustness check, I provide two alternative sets of results based on a limited set of control routes. First, I limit the routes to only those where the home tract is waterfront (Table 3, Column 6), this reduces the sample by 87%. Second, I provide results where the sample of routes is limited to only those where the home tract received a ferry terminal by the end of the study period (Column 7). This method essentially compares ferry-treated routes, to routes where the home neighborhood is treated by a ferry, but a ferry does not provide access to the work destination. The second limitation reduces the original sample by 98.6%. The alternative estimates range from 2.3-2.5, which are almost identical to the full sample specification (Column 1). While the full set of routes is large, the meaningful treatment variation is confined to a small number of routes. The high-level fixed effects mean the inclusion of irrelevant control routes does not significantly affect results.

I drop tracts between 200 and 1,000 meters from a ferry terminal as shown in Figure 4. Omitting these tracts might undercount the overall benefits of the ferry system by ignoring the partially treated areas. In Table 3, Column 8 I show results from an alternative TWFE specification that includes these intermediate tracts as a discrete treatment group. I find evidence of an increase in commuter flows to and from these areas, but the effect is small relative to the effect on tracts within 200 meters of a terminal.[16]

Table 4 breaks out results by the age of workers. LODES divides workers into those under 30 years of age, those 30-54, and those 55 or older. I refer to these groups as young, middle-aged, and older in analysis. I find that middle-aged workers are most responsive to the introduction of ferry service. Relative to pre-ferry levels, a ferry connection increased the flow of young workers by 18%, the flow of middle-aged workers by 52%, and the flow of older workers by 31%. In addition to appealing to higher-income workers, the ferry service also seemed to appeal disproportionately to middle-aged workers.

Table 4: Effect of Ferry Connection on Commuter Flow, By Age

	Young (1)	Middle-aged (2)	Older (3)
Ferry link dummy	0.302*	1.795**	0.310**
	(0.134)	(0.290)	(0.076)
Year fixed effects	Y	Y	Y
Home-work pair fixed effects	Y	Y	Y
N	82,294,092	82,294,092	82,294,092
Pre-treatment average	1.708	3.466	1.007
% change	+17.7%	+51.8%	+30.8%

Significance levels: $*$: 5% $**$: 1%. The estimation approach follows Callaway and Sant'Anna (2021). Standard errors are shown in parentheses.

In addition to testing for the effect of a ferry connection on route-level flows, I also test for induced changes in a tract's worker population using the tract-level worker counts described in the previous section (Equation 2). Table 5 estimates the partial effect of gaining ferry service on the number of workers living in those tracts, broken out by worker characteristics. I find a large increase in the total number of workers living within ferry-treated tracts. Gaining ferry service is correlated with an increase

[16] In Appendix B, I provide the accompanying event study analysis. I also reestimate the quantitative spatial equilibrium model under this alternative setup and provide estimated aggregate employment effects.

of 408 local workers. The overall increase in workers suggests that the ferry system was accompanied by an expansion in local housing supply. Schreurs et al. (2023) provides discussion of how ferry terminals were often accompanied by property development, particularly the creation of new condominium units close to terminals. A shift of original local workers into the labor force could also explain some of the rise in workers.

Table 5: Effect of Ferry Connection on Local Worker Population

	All Workers (1)	Low-income (2)	High-income (3)	F.I.R.E.† (4)	Professional Services (5)	Health Services (6)	Accom./ Food (7)
Ferry terminal dummy	407.818**	34.946	372.872**	85.903**	90.976**	13.699	18.680*
	(118.121)	(25.483)	(97.206)	(22.222)	(22.706)	(17.524)	(7.602)
Year fixed effects	Y	Y	Y	Y	Y	Y	Y
Tract fixed effects	Y	Y	Y	Y	Y	Y	Y
N	38,898	38,898	38,898	38,898	38,898	38,898	38,898
Pre-treatment average	1800.557	709.445	1091.112	298.504	233.995	243.693	121.936
% change	+22.6%	+4.9%	+34.2%	+28.8%	+38.9%	+5.6%	+15.3%

Significance levels: $*$: 5% $**$: 1%. The estimation approach follows Callaway and Sant'Anna (2021). Standard errors are shown in parentheses. †Finance, Insurance, and Real Estate.

The importance of Wall Street as a node of the ferry system may appeal disproportionately to workers in finance or professional services, industries that are concentrated in the Wall Street area. Additionally, many of the residential areas served by ferries had high home prices, which may also limit the ability of workers in lower-income industries from moving to benefit from ferry service. While I do not have detailed industry breakouts of commuter flows, I do have industry details for tract-level workforce counts. I test whether a ferry terminal changed the industry composition of the locally residing labor force (Table 5).

For the average treated tract, among the 408 new local workers, 373 (91%) were middle-to-high-income workers. I find significant increases in workers employed in FIRE industries (Finance, Insurance, and Real Estate) and Professional Services, with an increase in local workers of 29% and 39% respectively. Overall, 43% of the worker population growth can be attributed to these industries. Contrastingly, workers employed in lower-skilled, lower-paid industries such as Health Services and Accommodation and Food Services saw small, only marginally significant increases of 6% and 15%. The workforce growth effects are consistent with the ferry terminals attracting primarily middle-to-high-income workers employed in high-wage industries such as finance.

While the route level analysis can identify changes specific to ferry treated routes, the tract-level analysis of local workers can only capture general changes in the areas around terminals. The areas receiving ferry terminals were correlated with other development activities, particularly new housing development in Brooklyn and Queens along the East River. The estimates here capture both the direct effect of the ferry and local development activity that might be spatially and temporally correlated with ferry terminal construction.

In the context of New York City, rents are not purely a result of market forces. A 2023 housing survey showed nearly 1 million of New York City's 2.1 million rental units were subject to rent stabilization restrictions (Preservation and Development, 2023). Such rules blunt local rent changes and may discourage household moves. The estimated changes in commuter flows and worker populations may understate the long-run effects of the ferry system if household location is more elastic over longer time periods.

In Table 6 I conduct the same exercise for the number of jobs located within the ferry-treated tracts. In general, I do not identify statistically significant changes in the number of jobs sited in ferry-treated tracts after the ferry system opens. The estimates have large standard errors.

Table 6: Effect of Ferry Connection on Local Job Counts

	All Workers (1)	Low-income (2)	High-income (3)	F.I.R.E.† (4)	Professional Services (5)	Health Services (6)	Accom./ Food (7)
Ferry terminal dummy	1160.881	485.342	675.539	-385.035	306.032	297.391	130.584**
	(890.968)	(317.431)	(599.869)	(583.690)	(167.160)	(434.942)	(43.132)
Year fixed effects	Y	Y	Y	Y	Y	Y	Y
Tract fixed effects	Y	Y	Y	Y	Y	Y	Y
N	38,898	38,898	38,898	38,898	38,898	38,898	38,898
Pre-treatment average	4762.076	2183.152	2578.924	1447.363	469.487	709.297	131.248
% change	+24.4%	+22.2%	+26.2%	-26.6%	+65.2%	+41.9%	+99.5%

Significance levels: $*$: 5% $**$: 1%. The estimation approach follows Callaway and Sant'Anna (2021). Standard errors are shown in parentheses. †Finance, Insurance, and Real Estate.

For both worker and job estimates (Tables 5 and 6) I use the number of workers or jobs in the tract as the outcome variable. Given the strong skew in worker and particularly job counts across tracts, using the log-transformed version of worker and job counts may improve inference. While the presence of zeros again presents a challenge, zeros are relatively uncommon in this setting as most tracts have at least one worker and one job of each type. In Appendix C, I repeat the analysis of Tables 5 and 6 but log-transform the dependent variable. For workers, I find consistent results. For jobs, I again estimate statistically insignificant effects.

5 Structural Estimation of Worker Response

The regression analysis shows that New York City's ferry expansion had a measurable impact on commuting patterns. However, the observed shift in commuter flows could reflect a mix of mechanisms: some workers may have entered the labor force, while others simply changed home or work locations.[17] To isolate the roles of sorting from employment responses and recover aggregate effects, I turn to a structural model. The model enables estimation of how much of the observed changes in commuting reflect changes in employment versus relocation.

The proposed model builds on revealed route-level commuter behavior before and after ferry expansion, estimating route-specific preference parameters for low- and higher-income workers. I then recover the implied value of ferry access, that is consistent with observed flow changes, and simulate how workers adjust on both the neighborhood and employment margins. This modeling approach allows me to estimate not just the aggregate effects of ferry service, but also how benefits are distributed across worker types and locations. By capturing general equilibrium effects, the model offers insight into the broader labor market consequences of transit improvements, including why such investments may have unequal effects across income groups. The method can be applied to new public transit infrastructure that affects a subset of the commute routes within a larger city or region, where commuting flow data is available for both the period before and after the infrastructure began operating.

Understanding the mechanisms generating the change in commute flows has important policy implications. The introduction of a new transit system reduces the costs of some commutes, which may help unemployed workers attain employment, as predicted by the spatial job search and spatial mismatch literatures. Also consistent with the regression results would be that the observed changes in commuter flows are entirely explained by worker and firm location sorting with no new employment created. Tyndall (2021) demonstrated that new transit infrastructure can actually lower aggregate employment in a city where workers have heterogeneous location preferences. If employed workers with a high labor supply elasticity are crowded out of accessible areas by employed workers with low labor supply elasticity the equilibrium number of employed workers can decline. This could occur if the transit amenity pushes up local housing costs in central areas, causing central-city gentrification and low-skill worker

[17] For example, Glaeser et al. (2008) demonstrated workers who value transit may move toward transit.

displacement. Given that expansion of labor market opportunity, particularly for low-skilled or low-income workers, is a common goal of transit investment, it is important to understand how to design transit systems to maximize the labor market benefits.

Below, I outline a quantitative spatial equilibrium model. By pooling data across multiple years I provide time-invariant estimates of route preference for both worker types that allow the model to match average flow data. The route preference parameters are analogous to route-by-worker-type fixed effects. The values reflect average route preferences across the study period, both pre- and post-ferry. I then apply the time-variant, ferry-induced commuter flow changes estimated in the preceding section. I recover ferry preference parameters that precisely predict the shift in commuter flows towards ferry-treated routes. Knowing these parameters and the distribution of workers across routes, I can estimate aggregate effects and a distribution of direct benefits.

5.1 Structural Model Setup

I propose a static model of worker choice that will take the following general form. The utility of a worker is represented by a Cobb-Douglas style function represented by equation 3.

$$U_{ijk} = (C + \rho_{s(i)F(jk)})^{\gamma_{s(i)}} H^{(1-\gamma_{s(i)})} \chi_{s(i)jk} \tag{3}$$

Workers derive utility from numeraire consumption (C) and the consumption of generic units of housing (H). The share of income a worker spends on housing is set by $1 - \gamma_{s(i)}$. i indexes the worker, j indexes the home tract, k indexes the work tract, $s(i)$ indexes the income level of worker i, and $F(jk)$ indexes whether tracts j and k are connected by a ferry.

$\rho_{s(i)F(jk)}$ takes a value of zero if tracts j and k are not connected by a ferry route ($F(jk) = 0$). If tracts j and k are connected ($F(jk) = \mathbb{1}$), $\rho_{s(i)\mathbb{1}}$ is the consumption premium associated with the benefits of a ferry connection. $\rho_{s(i)\mathbb{1}}$ will be endogenously determined. Because $\rho_{s(i)F(jk)}$ enters additively with numeraire consumption (C) it will be interpretable as the worker's valuation of ferry service, expressed in dollars. Workers can be higher-income ($s(i) = h$) or low-income ($s(i) = l$) and this characteristic is fixed. $\chi_{s(i)jk}$ is a route-by-worker-type specific preference parameter. Some routes may provide higher utility than others based on their unique characteristics such as commuting time, traffic conditions, commuting mode options, housing and job prospects at the origin and destination tract, or any other route-specific, but time-invariant, characteristic. All workers of the same income type (low or higher) share a common evaluation over routes ($\chi_{s(i)jk}$) and this preference vector is time-invariant.

Each worker operates under a budget constraint (Equation 4). $w_{s(i)k}$ represents the wages paid to worker i. Employed workers earn a set wage dependent only on their type. I allow for non-employed workers by including a null element of work location choice ($k = \emptyset$). When a worker makes this selection they receive a reduced, but strictly positive, public assistance wage ($w_{i\emptyset} = w^P$). The variable p_j represents the price for a generic unit of housing in tract j. Workers choose housing quantity (H) consistent with utility maximization. All workers are renters and pay rent to a landlord outside of the local economy.

$$w_{s(i)k} = Hp_j + C \tag{4}$$

I combine the utility function and budget constraint to generate an indirect utility function (Equation 5). ξ_{ijk} follows a Type 1 extreme value distribution and captures a worker's individual idiosyncratic preferences over each commuting route. I follow the literature by introducing a shape parameter (θ) to capture the degree to which workers are willing to substitute across discrete choices.

$$\begin{gathered} V_{ijk} = \theta \ln(w_{s(i)k} + \rho_{s(i)F(jk)}) - \theta(1 - \gamma_{s(i)})\ln(p_j) + \chi_{s(i)jk} + \xi_{ijk} \\ V_{ijk} \equiv v_{ijk} + \xi_{ijk} \end{gathered} \tag{5}$$

The extreme value distributed idiosyncratic error term produces a multinomial logit probability function (Equation 6), capturing the probability a worker selects a specific home-work pair (P_{ijk}). The upper bar notation indicates the maximum value in the set.

$$P_{ijk} = \frac{e^{v_{ijk}}}{\sum_1^{\bar{j}} \sum_1^{\bar{k}} e^{v_{ijk}}} \tag{6}$$

I model firms such that each tract contains a single representative profit-maximizing firm. Each firm (tract) has a production technology that takes low-income labor, higher-income labor, and capital as inputs. Firms exist in a competitive market, possess constant returns to scale production technology, have access to a perfectly elastic external capital market, and earn zero profits. In such an environment, firms will expand to hire as many workers as are willing to accept employment at a persistent wage level. Therefore, worker wages vary by worker type, and by whether the worker is employed, but are otherwise fixed in the model. Firm wage offers do not vary across space. Any qualitative differences across firms are subsumed by the route level preference parameters. One reason low (or higher) income workers may prefer working at a particular location is the job characteristics of that local firm, but this effect is not uniquely identified relative to any other attractive quality of that route.

I describe a static model. When estimating the impacts of the ferry system I reestimate the model under different ferry configurations to recover how expansions of the ferry system impact the distribution of worker route choices.

5.2 Structural Model Solution Method

To solve the model I impose eight exogenous structural parameters (Table 7). The annual wages for low and higher-income workers are set to \$18,000 and \$69,000 respectively. I recover these estimates from the 2013 5-year ACS data, which provides average earnings for workers earning above and below the \$40,000 income group cut-off as defined in the LODES data. I set non-employment assistance income (w^P) equal to \$4,300, which is the median personal income of a non-working New York City resident, which includes social assistance income, according to the 2013 5-year ACS microdata from IPUMS. I parameterize the share of income spent on housing (γ) with the same IPUMS microdata. The median worker earning above \$40,000 spends 17% of their income on either rent or mortgage payments. Among those earning below \$40,000, microdata indicates the median worker spends 54% of their income on housing, reflecting the expensive housing market of New York City and the low cutoff used to define the low-income population. I therefore set $\gamma_{s(i)=l} = 0.46$ and $\gamma_{s(i)=h} = 0.83$. I impose a shape parameter (θ) value of 6.4, following (Ahlfeldt et al., 2015). The estimate represents a moderate value as the literature has estimated lower values in some settings (Severen, 2023; Tsivanidis, 2023) and higher values in others (Kreindler and Miyauchi, 2023).

When ferry connections are introduced, I limit the model's solution space to match the estimated changes in commuter flows identified for low and higher-income workers in Section 4.2. I found that a ferry connection increased the average number of low and higher-income workers commuting on a route by 0.296 and 4.882 respectively (Figure 5).[18] The model solution generates these shifts by identifying the values workers place

[18] I assume the final estimate of the event study represents the long-run effect of a ferry connection. As an alternative method, I estimate the model using the difference-in-difference point estimates from

Table 7: Exogenous Model Parameters

Symbol	Value	Description
$w_{s(i)=l}$	18	Annual income for low-income workers ($1,000s)
$w_{s(i)=h}$	69	Annual income for higher-income workers ($1,000s)
w^P	4.3	Annual public assistance income for non-employed workers ($1,000s)
$\gamma_{s(i)=l}$	0.46	Share of income spent on non-housing consumption for low-income workers
$\gamma_{s(i)=h}$	0.83	Share of income spent on non-housing consumption for higher-income workers
$F_{s(i)=l}$	0.296	Ferry treatment effect on low-income commuter flow
$F_{s(i)=h}$	4.882	Ferry treatment effect on higher-income commuter flow
θ	6.8	Substitutability shape parameter

I impose eight parameters on the model. Income and housing expenditure estimates are taken from the 2013 5-year ACS.

on a ferry connection ($\rho_{l\mathbb{1}}$, $\rho_{h\mathbb{1}}$) that generate worker flow shifts equal to those estimated in the regression analysis.

In addition to the exogenous parameters shown in Table 7, I impose tract-level rents for the pre-ferry period but allow rents to adjust endogenously in subsequent periods. I parameterize pre-ferry rents using data from the 2013 5-year ACS. For the tracts where rent data is not available, I impute the average city-wide rent ($14,400 annually). When a worker selects to live in tract j they pay the median market rent for that tract (p_j), multiplied by their housing quantity choice H. Rents are allowed to endogenously adjust to clear the housing market of each tract once ferry service is introduced. In reality, a significant share of residents are insulated from rent changes by rent control policies and public housing. The model does not account for this market rigidity and assumes residents face market rents. The initial rent levels do not have a first-order importance because variation is subsumed by the route level preference parameters ($\chi_{s(i)jk}$) in estimation.

Most quantitative spatial models do not consider unemployment (Redding and Rossi-Hansberg, 2017). To allow for workers who are not employed, I add one route choice for every populated tract, where the work location is null. I use 2011 ACS data for the share of the over-16 population in each tract that is not in the labor force. I assign every tract a population of low and higher-income workers who are not employed to match the overall rate in the ACS data. While the LODES data supplies the number of employed workers completing each route, I add in additional workers

Table 3, which were 0.243 and 2.165. Adopting these values generally lowers the magnitude of structural model estimates. For example, I find the ferry consumption premiums are 24% lower for low-income workers and 45% lower for higher-income workers. The estimate for new employment gained among higher-income workers falls by half, and the loss of low-income worker employment becomes close to zero.

for the non-employed. The proposed approach incorporates unemployment without requiring a nested labor supply decision or requiring the modeling of search frictions (for an example with search frictions see Pérez Pérez (2022)). Because worker preference varies at the route level, a worker faces distinct utilities from being unemployed in any particular home tract. The method is analogous to adding one new work destination that can be reached (costlessly) from every unique home location. The modeling choice implies a particular worker behavior wherein substitution into unemployment is decided in the same way as substitution to a different home-work route. The approach simplifies the model but imposes a specific substitution behavior.

For the 2002-2019 study period, the level of ferry service can be effectively divided into four periods (Figure 2). The pre-system period spans 2002-2010. The second period (2011-2016) includes only ferry service for the East River route. The third period (2017) adds service for the Rockaway, South Brooklyn, and Astoria routes. The fourth period (2018-2019) includes service for all six routes.

I proceed to solve the model in two steps. First, I estimate the vector of route-by-worker-type consumption premiums, $\chi_{s(i)jk}$. I collapse the 18 years of data to the route-by-worker-type level, taking the average values of bilateral commuter flows across all years. I set rents and wages according to the exogenous values. I then use contraction mapping to recover the unique vector $\chi_{s(i)jk}$ that generates the observed, average bilateral commuter flows. In solving for the equilibrium, elements of $\chi_{s(i)jk}$ are raised or lowered to precisely match the averaged commute flow data. Preference parameters are identified using cross-sectional variation. For example, two routes might provide the same observed utility, but commuter flows reveal one of these routes to be much more popular. I account for this difference in popularity by raising the relative value of that route's consumption premium ($\chi_{s(i)jk}$).

By pooling the data, I avoid relying on an overly sparse matrix. Using the same LODES data from New York City, Dingel and Tintelnot (2020) argued that a single year of data produced severe bias from overfitting while pooling three years of data significantly reduced the issue. In the method proposed in this paper, I pool 18 years of data, which significantly reduces the sparseness of the matrix and reduces concerns about overfitting the data. A primary concern in Dingel and Tintelnot (2020) is that out-of-sample counterfactual estimates will be biased by the significant amount of idiosyncratic behavior captured in the granular data. In my primary analysis, I do not estimate a purely hypothetical scenario, rather I constrain the model to generate commuter flow changes that were produced by an event with observable effects.

Of 9.2 million possible routes and worker types in the cross-sectional matrix, 53% have zero commuters recorded for every year of the LODES data. In effect, the model assigns a χ_{jk} for these routes equal to negative infinity. It is computationally equivalent to drop these routes from the analysis. I observe the post-treatment period and therefore know that, even after ferry service was available, these routes still attracted no commuters, suggesting these routes are unlikely to correspond to viable routes even in the counterfactual environment.

In solving $\chi_{s(i)jk}$, I produce two vectors of route-level preference parameters, one for each worker-type. Tract pairs that provide easy commutes will be more popular (conditional on differences in rents in the home tract) and therefore require a higher value of $\chi_{s(i)jk}$ to match the data. Estimating $\chi_{s(i)jk}$ averaged over the entire study period allows for the vector to be interpreted as the time-invariant component of common route preferences.

Moving to the second step of the estimation method, I hold constant the route-level consumption premiums ($\chi_{s(i)jk}$) that were generated in the first step and expand the data to include the four distinct periods of ferry service. Holding route preferences ($\chi_{s(i)jk}$) constant, I limit the model to generate the route-level ferry treatment effects calculated in the regression analysis (F^l and F^h). I allow the worker-type specific consumption premiums for ferry service ($\rho_{s(i)=l,F(jk)=\mathbb{1}}$, $\rho_{s(i)=h,F(jk)=\mathbb{1}}$) to adjust to generate the correct change in commuter flow. For example, $\rho_{h\mathbb{1}}$ will adjust so that a ferry connection induces an average of 4.882 additional higher-income workers to select a ferry-serviced route relative to the average change in workers selecting non-ferry-serviced routes. I allow the vectors of period-specific rents to adjust to clear the housing market by attracting the correct number of residents to each tract, though a tract's mix of low and higher-income workers may change.

Positive ferry amenity values (F^l and F^h) will push up demand to live in tracts that are connected by a ferry route. The increase in demand, and fixed housing supply, results in an increase in rents. The increase in rents affects both low and higher-income workers as they compete in a common housing market.

I define the model solution of the second step as a vector of rents and ferry preference parameters where each worker chooses the commuting route that maximizes their utility, the housing market clears, and the shifts in commuting flows towards ferry routes match the observed values (F^l and F^h). The conditions represent a Nash equilibrium for all workers.

Spatial sorting models with agglomeration economies, or other interactions across agents, can give rise to multiple possible equilibria. Workers in the above model do not have a preference for the characteristics (eg income) of their neighbors. As demonstrated in Bayer and Timmins (2005), uniqueness is achieved when there are positive congestion effects and agglomeration economies are sufficiently small. The model I propose has congestion costs (endogenous rents) but no agglomeration economies, which is sufficient to establish uniqueness.

5.3 Structural Model Results

The main parameters of interest for the structural estimation model are the consumption premiums associated with route ferry access for both low and higher-income workers ($\rho_{s(i)=l,F(jk)=\mathbb{1}}$, $\rho_{s(i)=h,F(jk)=\mathbb{1}}$). $\rho_{s(i)F(jk)}$ takes a value of zero for tract pairs (jk) that are not connected by a ferry, by construction. In the equilibrium solution, for tract pairs connected by a ferry route, $\rho_{s(i)\mathbb{1}}$ takes a value of 0.516 for low-income workers and 5.360 for higher-income workers (Table 8). The values can be interpreted in the model as consumption premiums represented in \$1,000s, meaning low and higher-income workers value a ferry connection as being equal to \$516 and \$5,360 of annual consumption respectively (Table 8). As demonstrated in the regression analysis, it is primarily higher-income workers who adjust their behavior to make use of the ferry. Because the ferry is rarely faster than the subway, these benefits are not necessarily due to predicted reductions in commute time but a premium for access to a more reliable and comfortable commute mode.

Table 8: Structural Model Solution Parameters

Symbol	Value	Description
$\rho_{l\mathbb{1}}$	0.516	Ferry connection consumption premium (\$1,000), low-income workers
$\rho_{h\mathbb{1}}$	5.360	Ferry connection consumption premium (\$1,000), higher-income workers

Solving the model yields the consumption premium parameters.

Wages for higher-income workers in the model are 3.8 times that of low-income workers. As a fraction of income, low-income workers value a ferry connection at 2.9% of income, while higher-income workers value the connection at 7.8% of income. Because of a higher valuation of a ferry connection, and a higher preference for ferry-connected routes, higher-income workers capture most benefits, as discussed below.

I interpret the values of $\rho_{s(i)\mathbb{1}}$ cautiously for two reasons. First, the methodology of recovering route level preference parameters using pooled data will be impacted by post-ferry observations, which may inflate estimated route preferences on treated routes and subsequently affect the interpretation of $\rho_{s(i)\mathbb{1}}$. Second, the estimates of $\rho_{s(i)\mathbb{1}}$ are influenced by exogenously imposed parameters. Different imposed parameters can effect the magnitude of $\rho_{s(i)\mathbb{1}}$ estimates.

In equilibrium, the provision of a local ferry connection increases local rents for treated tracts, as it represents a local amenity. The results of the structural model suggest a ferry terminal causes an average annual rent increase of \$302 (\$25 per month) or a 1.5% increase relative to the average pre-ferry rent in treated tracts. The increase in rent makes the tract less desirable, ceteris-paribus, repelling enough residents so the local housing market clears.

A positive rent effect is consistent with reduced form evidence. A government report on the effects of the initial East River Ferry line found an increase in local residential property values around terminals (NYCEDC, 2013). I test for a rent effect using 5-year ACS estimates of tract-level rents, centered at the middle year of the survey period. I use a regression approach analogous to Equation 2. Data limitations require the analysis to span only 2007-2019. I estimate that a local ferry connection is correlated with an average increase in monthly rent of \$418, or 26% of the pre-ferry treated tract median. The estimate confirms a positive rent effect, but the estimated magnitude is far larger than what is implied by the model. While the structural model recovers the direct price effects of the ferry from increased local demand for housing, the ACS estimates will also capture upgrades to housing stock. Increases in housing quality accompanying local upzoning and gentrification could explain the large effect estimated in ACS data.

The route-flow analysis suggested that few low-income workers are incentivized to switch to a commuting route with a ferry (Table 3). If low-income workers placed no value on a ferry connection ($\rho_{l\mathbb{1}} = 0$) they would still be exposed to the rent increases around ferry terminals brought on by the behavior of higher-income workers. If low-income workers did not value the ferry at all, we would expect a decline in the number of low-income workers commuting along ferry routes, as they are repelled from living in ferry terminal areas as local rents increase. The positive valuation of ferry service estimated for low-income workers (\$516 annually) more than offsets the repellent effect of higher rents, leading to a marginal increase in low-income commuters on ferry-connected routes.

Solving the initial pooled version of the model recovers preferences over routes. If route-specific preferences were weak, workers would all want to live in the tract where rent is lowest. The revealed preferences of workers show a willingness to bear higher rents for a neighborhood that they prefer. While the benefits of the ferry are large relative to the rent change, few workers are willing to alter their home, work, or labor market participation decision to capture this surplus. This is because they are unlikely to have personal preferences that align with where the ferry is servicing. By moving, they sacrifice the utility gained by being in the neighborhood they would otherwise prefer. The reason higher-income workers are more likely to switch routes is that their route-specific preferences are more likely to align with the ferry routes.

The model solution reveals that ferry routes align with the route-specific preferences of higher-income workers more so than low-income workers. Table 9, column 1 tests for a correlation between route preference parameters ($\chi_{s(i)jk}$) and whether the route was selected for ferry service. I standardize the route preference parameters for each worker type to be mean zero and standard deviation equal to one. I look across routes with at least one commuter of that type, as these have defined preference parameters. Column 1 shows the result of a cross-sectional regression of $\chi_{s(i)jk}$ against a dummy variable for whether that route had a ferry connection by the end of the study period. Routes treated by a ferry had a preference parameter 1.43 standard deviations above the mean for low-income workers, but 1.74 standard deviations above the mean for higher-income workers. Because ferry-serviced areas are generally central, high-amenity areas, both groups have a preference to live or work along ferry-serviced routes, relative to the average city route, and are willing to bear higher rents to live there. However, higher-income workers have a stronger bias towards these routes.

Table 9: Correlation of Ferry Treatment Status and Recovered Route-level Preference Parameters ($\chi_{s(i)jk}$)

	Route Level (1)	Home Tract (2)	Work Tract (3)
		Low-income Workers	
Ferry link dummy	1.433**	0.846**	0.297
	(0.065)	(0.194)	(0.227)
N	2,467,060	2,124	2,158
		High-income Workers	
Ferry link dummy	1.737**	1.436**	0.882**
	(0.082)	(0.221)	(0.248)
N	1,823,702	2,124	2,160

Significance levels: $*$: 5% $**$: 1%. Robust standard errors shown in parentheses.

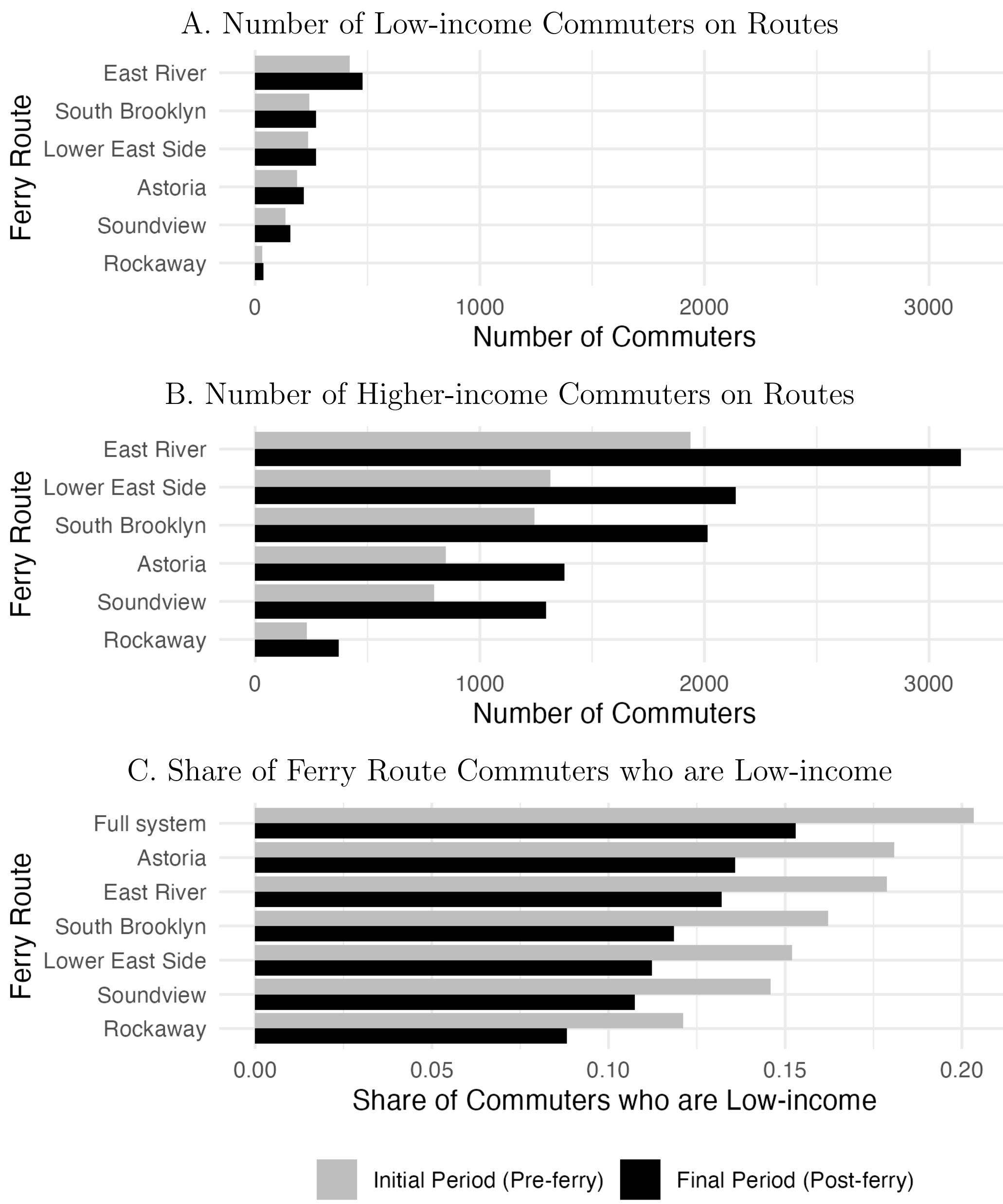
I subsequently break out the differences in preference by home and work tracts. I take the weighted average of the preference parameters across routes, collapsed to the tract. When calculating average home tract preference I weight by the number of residents, and for work tract preference I weight by number of jobs. I recover the average route preference for every home and work tract, completing separate calculations for low and higher-income workers. I regress the normalized tract preference level against whether that tract was provided a ferry terminal by the end of the study period. For preference over home tracts (column 2), I find that both low and higher-income workers are willing to pay higher than average rents to live in the tracts that gained a ferry terminal, but the preference is stronger for higher-income workers. Having a ferry connection is correlated with a 0.85 standard deviation increase in low-income worker home location preference but a 1.44 standard deviation increase for higher-income workers. For work location preference (column 3), I find low-income workers have no preference towards working in ferry-serviced tracts, whereas higher-income workers are significantly more amenable to working in ferry-serviced tracts. The correlation between revealed location preference parameters and ferry service confirms that ferry locations were directed to places where higher-income workers prefer to live and, in particular, prefer to work.

Table 9 results arise from ferry terminals being located in neighborhoods popular among higher-income residents (eg Williamsburg) and popular work destinations for higher-income workers (eg Wall Street). Location preferences make higher-income workers much more willing to substitute towards ferry-serviced routes relative to low-income workers.

The model generates full worker distributions across routes and I can recover pre- and post-ferry worker distributions. Figure 6A shows the total number of low-income workers commuting on each ferry-serviced route, contrasting the start and end of the study period. Figure 6B provides results for higher-income workers, and Figure 6C shows the ratio of low to higher-income workers on each route. While a connection generates an increase in flows for both worker types along all routes, the effect is far larger for higher-income workers. The Astoria route covered the highest initial share of low-income workers (18%). However, this also resulted in a relatively high number of low-income workers being exposed to rising rents.

The results of the model also yield shifts in the share of workers who are employed relative to those out of the workforce. After the ferry service is fully implemented, the model solution indicates 737 higher-income workers moving into the labor force and 37 low-income workers exiting the labor force. Across the entire workforce, the change amounts to a rise in the city-wide employment rate of 0.02%. I estimated that 2,600 new

Figure 6: Commuters on Each Ferry Route

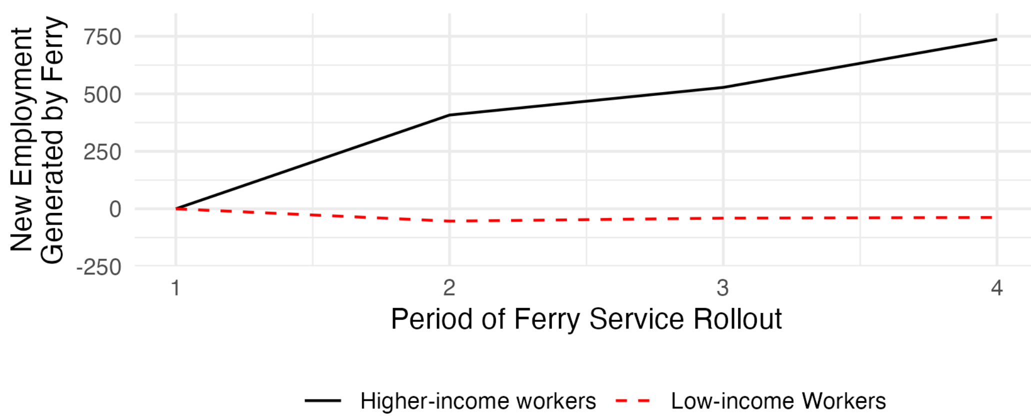


The bars indicate the number or share of workers who complete a commute that aligns with each of the NYC Ferry routes. Some commute routes are served by multiple ferry routes, meaning some commuters are counted for multiple routes. In panel C, the share of low-income workers across the full system exceeds that of any individual route because higher-income workers are more common on routes with multiple ferry connections.

labor market connections were generated by the ferry system. Therefore, reconciling the regression and structural results implies that 27% of the new labor market connections were the result of workers entering the labor force because of the transit improvement, while 73% of new connections were from already employed workers altering either their home or work location to benefit from a ferry commute. The model solution implies that ferry service actually reduced city-wide low-income worker employment, in contrast with stated policy goals.

Figure 7 graphs the number of workers who became employed due to the availability of ferry service. The figure shows changes as the ferry routes are rolled out in phases. For higher-income workers, each expansion raises aggregate employment, with the initial expansion (The East River Ferry route) having the largest effect. The initial East River Ferry route marginally reduced low-income employment, as low-income workers had low preferences for the terminals of this route, and local rent increases repelled low-income workers from the central neighborhoods serviced by the route, displacing them to more isolated areas where they are more likely to exit employment. The subsequent route expansions had a negligible effect on low-income worker employment.

Figure 7: Number of Workers who Became Employed Due to Ferry Service



Model results deliver estimated changes in aggregate employment. Period one corresponds to no ferry service, and subsequent periods represent each discrete expansion of ferry service, as recorded in Figure 2. When all six routes are operating, the new equilibrium implies a net employment increase of 700 workers.

In the final equilibrium, there were 951 low-income workers and 5,268 higher-income workers with commutes between tract pairs connected by NYC Ferry. While

56% of the workforce is in the low-income group, only 15% of commuters on ferry-serviced routes are low-income. Using the estimated annual benefits of access to a ferry route ($\rho_{s(i)\mathbb{1}}$), I estimate that the ferry service provides \$491,000 in annual benefits to low-income workers who can commute directly through the ferry system, compared to \$28 million for higher-income workers. This only accounts for benefits among those whose home and work census tracts are directly linked through the ferry service. These values are subject to the uncertainty in the estimates of $\rho_{s(i)\mathbb{1}}$, as outlined above. The methodology also ignore potential spillover effects, where the commute benefits of ferries extend beyond the 200 meter treatment radius. In Appendix B, I provide an alternative analysis that extends the treatment radius. The alternative specification finds the spillover effects are generally small. Despite these limitations, the estimates give an idea of the distribution of direct commuting benefits, with 98% of benefits accruing to the higher-income group. These direct benefit estimates are small in comparison to the annual cost of providing ferry service, which was estimated to be \$105 million.

Understanding the impacts of the individual routes can hold policy lessons for route selection. In Appendix D, I provide estimates on the local employment effects of each route (Figure D1). I also provide full structural model employment estimates under scenarios where I "turn on" one route at a time (Figure D2). I find all six routes cause a positive aggregate employment effect for higher-income workers and a marginally negative employment effect for low-income workers. Servicing routes with few low-income workers (Soundview, Rockaway) has almost no effect on low-income employment while raising higher-income employment, whereas routes with a relatively high share of low-income workers to begin with (East River) triggers larger low-income job losses.

The divergent aggregate employment outcome for the two income groups is a result of the routes selected. In Appendix E, I provide results for a hypothetical route that serves tract pairs where low-income workers hold relatively high preference parameters. Commutes from Coney Island and East New York to Wall Street were found to fit this description. I provide a hypothetical route map in Figure E1 that illustrates these connections. Using the model's route and ferry preference parameters, and resolving the model under this alternative route yields positive aggregate employment effects for both groups (Figure E2). The exercise demonstrates how the methodology of this paper could be used to compare prospective routes in terms of employment outcomes and the distribution of benefits.

6 Conclusion

Estimating the impact of a new transit system is complicated by endogenous worker decisions. I provide a method for estimating the impact of New York City's ferry service expansion on the local workforce. I distill lessons from recent structural models that deal with new transit infrastructure and panel bilateral commuter flow data (Tsivanidis, 2023; Severen, 2023; Dingel and Tintelnot, 2020; Tyndall, 2021). I propose a simplified model that has limited data requirements and a parsimonious solution strategy. The use of structural estimation models in similar scenarios could provide valuable insight into the impacts of new transit systems. Applying the model to NYC's ferry system demonstrates that small but significant labor market improvements can be attributed to the system.

Detailed commuter flow data is available across the US through the LEHD LODES as well as the Census Transportation Planning Products (CTPP). LODES data is now available for 21 years, with annual updates continuing. I leverage multiple years of data to estimate route-by-worker-type preference parameters. The issue of sparseness in large commuting matrices continues to be a topic of econometric concern in spatial structural estimation of commuting flows. I propose a pooling method that greatly reduces matrix sparseness. I contribute to the development of empirical methods that leverage longitudinal variation to estimate the total impacts of new transit infrastructure. Deriving reasonable parameters of worker preferences for transportation links could also inform future analyses that estimate the impact of hypothetical future transportation systems. Recovering a full matrix of commuter preference parameters could allow infrastructure planning to target benefits at particular populations by providing infrastructure on routes they are likely to use.

A limitation of estimating route preferences with multiple years of pooled data is it requires an assumption that worker preferences across routes are stable over time. As neighborhoods evolve, the spatial preferences of workers will change. Estimating preferences on pooled data implies a trade-off between avoiding matrix sparseness and introducing data that may not reflect current preferences.

The effectiveness of NYC Ferry was the topic of significant policy debate. Of interest was whether the service improved labor market connections within the city and whether these benefits accrued to low or high-income workers. I provide some answers to these policy questions. I find the ferry service had a statistically significant but small effect on commuter flows within the city. I estimate roughly 2,600 workers altered their behavior to take advantage of the reduced commuting costs offered by the ferry service. All net commuting growth was among workers with incomes of $40,000 or higher. Results from the structural model show low- and higher-income workers both value a ferry connection. Initially, higher-income workers were overrepresented on routes selected for ferry service, causing them to capture most of the benefit. Additionally, the location preferences of higher-income workers aligned more closely with ferry routes, resulting in a shift of higher-income workers towards ferry-serviced routes. I find no evidence that the ferry system was successful in expanding low-income employment.

Few studies directly estimate aggregate employment effects of new transit. For the case of light rail, Tyndall (2021) found new US light rail systems marginally decreased metropolitan employment, with every 10 new stations generating a 0.6 percentage point drop in the metro-wide employment rate, with losses concentrated among those with low earning potential and the effects driven by residential displacement. For BRT in Bogota, Tsivanidis (2023) estimated positive labor market effects across low and high income earners, but found larger benefits accruing to high earners. The NYC Ferry system appears to similarly direct benefits towards higher income households. Modeling residential sorting and rent impacts are important to uncovering the distributional consequences of new public transit infrastructure.

A limitation of this study is its narrow focus on direct commuting benefits. While commuter benefits were a primary motivation for the construction of the system, the ferry may be providing numerous other economic and social benefits that I do not account for. The real estate development triggered near stations represents a significant source of economic value, which could also increase property tax revenue for the city. Economic activity from tourism could increase, as the ferry provides a scenic and enjoyable trip. Non-work trips for local residents are also unaccounted for. Nearly 60% of trips are non-commuting trips (New York City Economic Development Corporation, 2019). To the extent ferry trips displace car trips, the shift could reduce local pollution, carbon emissions, and other externalities from vehicles. While I provide evidence that the commuting benefits of the ferry were far less than system costs, accounting for other benefits could demonstrate the system's value.

The high cost of transportation infrastructure in the US (Brooks and Liscow, 2022) means that recent transit projects often have costs exceeding their estimated direct benefits. However, I find the direct benefits of NYC Ferry are particularly small relative to system costs. For comparison, Gupta et al. (2022) analyzed the real estate value uplift caused by New York City's Second Avenue Subway extension, finding $5.5

billion in new real estate value created, which exceeded the \$4.5 billion in construction costs. The estimates were relatively consistent with estimates of the value of travel time-saving. The ferry system provides a different setting, where scheduled commute-time savings are minimal but significant benefits may accrue from improved trip comfort and reliability.

Commuting route preference differences across income groups are important determinants of who is served by a particular piece of transportation infrastructure. Designing equitable transit systems requires spatially targeting infrastructure to routes where there is a demand among the workers who are intended to benefit from the system.

References

Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., and Wolf, N. (2015). The economics of density: Evidence from the Berlin wall. *Econometrica*, 83(6):2127–2189.

Alonso, W. (1964). *Location and land use. Toward a general theory of land rent.* Cambridge. Harvard University Press.

Anas, A. (1981). The estimation of multinomial logit models of joint location and travel mode choice from aggregated data. *Journal of Regional Science*, 21(2):223–242.

Andersson, F., Haltiwanger, J. C., Kutzbach, M. J., Pollakowski, H. O., and Weinberg, D. H. (2018). Job displacement and the duration of joblessness: The role of spatial mismatch. *Review of Economics and Statistics*, 100(2):203–218.

Bayer, P., Ferreira, F., and McMillan, R. (2007). A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy*, 115(4):588–638.

Bayer, P. and McMillan, R. (2012). Tiebout sorting and neighborhood stratification. *Journal of Public Economics*, 96(11-12):1129–1143.

Bayer, P., McMillan, R., and Rueben, K. (2004). An equilibrium model of sorting in an urban housing market. *National Bureau of Economic Research.*

Bayer, P. and Timmins, C. (2005). On the equilibrium properties of locational sorting models. *Journal of Urban Economics*, 57(3):462–477.

Brooks, L. and Liscow, Z. (2022). Infrastructure costs. *American Economic Journal: Applied Economics.*

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

Campion, S. (2019). Swimming in subsidies: The high cost of NYC ferry. *Citizens Budget Commission.*

Chernoff, A. and Craig, A. N. (2022). Distributional and housing price effects from public transit investment: Evidence from Vancouver. *International Economic Review*, 63(1):475–509.

Dingel, J. I. and Tintelnot, F. (2020). Spatial economics for granular settings. *National Bureau of Economic Research.*

Epple, D. and Sieg, H. (1999). Estimating equilibrium models of local jurisdictions. *Journal of Political Economy*, 107(4):645–681.

Ferreyra, M. M. (2007). Estimating the effects of private school vouchers in multidistrict economies. *The American Economic Review*, 97(3):789–817.

Fujita, M. and Ogawa, H. (1982). Multiple equilibria and structural transition of non-monocentric urban configurations. *Regional Science and Urban Economics*, 12(2):161–196.

Glaeser, E. L., Kahn, M. E., and Rappaport, J. (2008). Why do the poor live in cities? The role of public transportation. *Journal of Urban Economics*, 63(1):1–24.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.

Gupta, A., Van Nieuwerburgh, S., and Kontokosta, C. (2022). Take the Q train: Value capture of public infrastructure projects. *Journal of Urban Economics*, 129:103422.

Holzer, H. J., Quigley, J. M., and Raphael, S. (2003). Public transit and the spatial distribution of minority employment: Evidence from a natural experiment. *Journal of Policy Analysis and Management*, 22(3):415–441.

Kain, J. F. (1968). Housing segregation, negro employment, and metropolitan decentralization. *The Quarterly Journal of Economics*, 82(2):175–197.

Kreindler, G. E. and Miyauchi, Y. (2023). Measuring commuting and economic activity inside cities with cell phone records. *Review of Economics and Statistics*, 105(4):899–909.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, pages 105–142.

Mills, E. S. (1967). An aggregative model of resource allocation in a metropolitan area. *The American Economic Review*, 57(2):197–210.

Mo, J. (2023). Heterogeneous effects of urban transport infrastructure on population distribution: The role of educational access. *Regional Science and Urban Economics*, 101:103902.

Muth, R. F. (1969). Cities and housing; the spatial pattern of urban residential land use.

New York City Economic Development Corporation (2019). NYC ferry 2019 summer survey.

NYC Ferry (2019). NYC Ferry Quarterly Update, Q4.

NYCEDC (2011). New York City Economic Development Corporation, Comprehensive Citywide Ferry Study.

NYCEDC (2013). New York City Economic Development Corporation, Citywide Ferry Study.

Pérez Pérez, J. (2022). City minimum wages and spatial equilibrium effects. Technical report, Working Papers.

Preservation, N. H. and Development (2023). New york city housing and vacancy survey, selected initial findings.

Redding, S. J. and Rossi-Hansberg, E. (2017). Quantitative spatial economics. *Annual Review of Economics*, 9(1):21–58.

Sandell, R. (2017). Network design strategies to increase efficiency and usefulness of urban transit ferry systems. *Transportation Research Record*, 2649(1):71–78.

Schreurs, G., Scheerlinck, K., and Gheysen, M. (2023). The local socio-economic impact of improved waterborne public transportation. the case of the New York City ferry service. *Journal of Urban Mobility*, 3:100042.

Severen, C. (2023). Commuting, labor, and housing market effects of mass transportation: Welfare and identification. *Review of Economics and Statistics*, 105(5):1073–1091.

Sieg, H., Smith, V. K., Banzhaf, H. S., and Walsh, R. (2004). Estimating the general equilibrium benefits of large changes in spatially delineated public goods. *International Economic Review*, 45(4):1047–1077.

Thompson, R., Burroughs, R., and Smythe, T. (2006). Exploring the connections between ferries and urban form: Some considerations before jumping on board. *Journal of Urban Technology*, 13(2):25–52.

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

Tsivanidis, N. (2023). Evaluating the impact of urban transit infrastructure: Evidence from Bogota's Transmilenio. *Working Paper*.

Tyndall, J. (2017). Waiting for the R train: Public transportation and employment. *Urban Studies*, 54(2):520–537.

Tyndall, J. (2018). Bus quality improvements and local commuter mode share. *Transportation Research Part A: Policy and Practice*, 113:173–183.

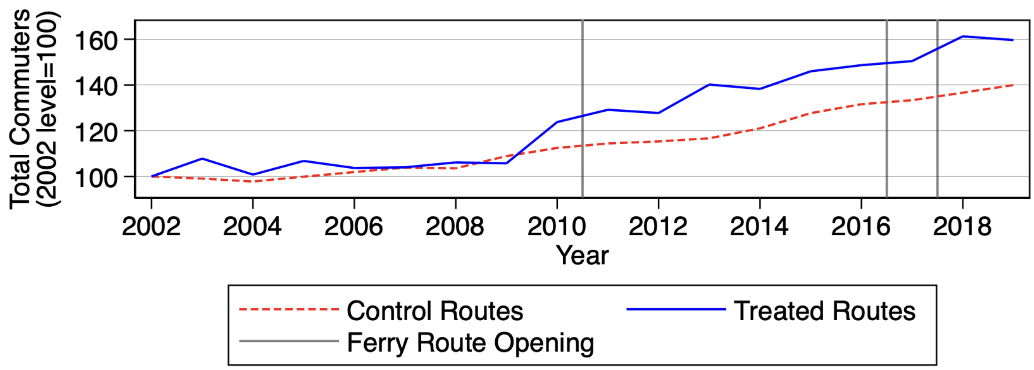
Tyndall, J. (2021). The local labour market effects of light rail transit. *Journal of Urban Economics*, 124:103350.

Vilain, P. B., Cox, J., and Mantero, V. (2012). Public policy objectives and urban transit: Case of passenger ferries in the New York City region. *Transportation research record*, 2274(1):184–191.

Appendix A

Figure A1 provides an unconditional parallel trends comparison. I divide all routes according to whether they were connected by a ferry route by the end of the study period, and compare the total number of commuters using those routes normalized to a 2002 baseline. I find commuter growth on treated and control routes was very similar from 2002-2009. While the first ferry route was opened in 2011, I find an increase in commuters on treated routes beginning in 2010, suggesting an anticipation effect wherein commuters may have adjusted home or work locations in anticipation of a ferry connection. However, this unconditional anticipation effect could be spurious as there is no apparent anticipation effect in the event study analysis (Figure 5), which incorporates control variables.

Figure A1: Parallel Trend of Treated and Control Routes

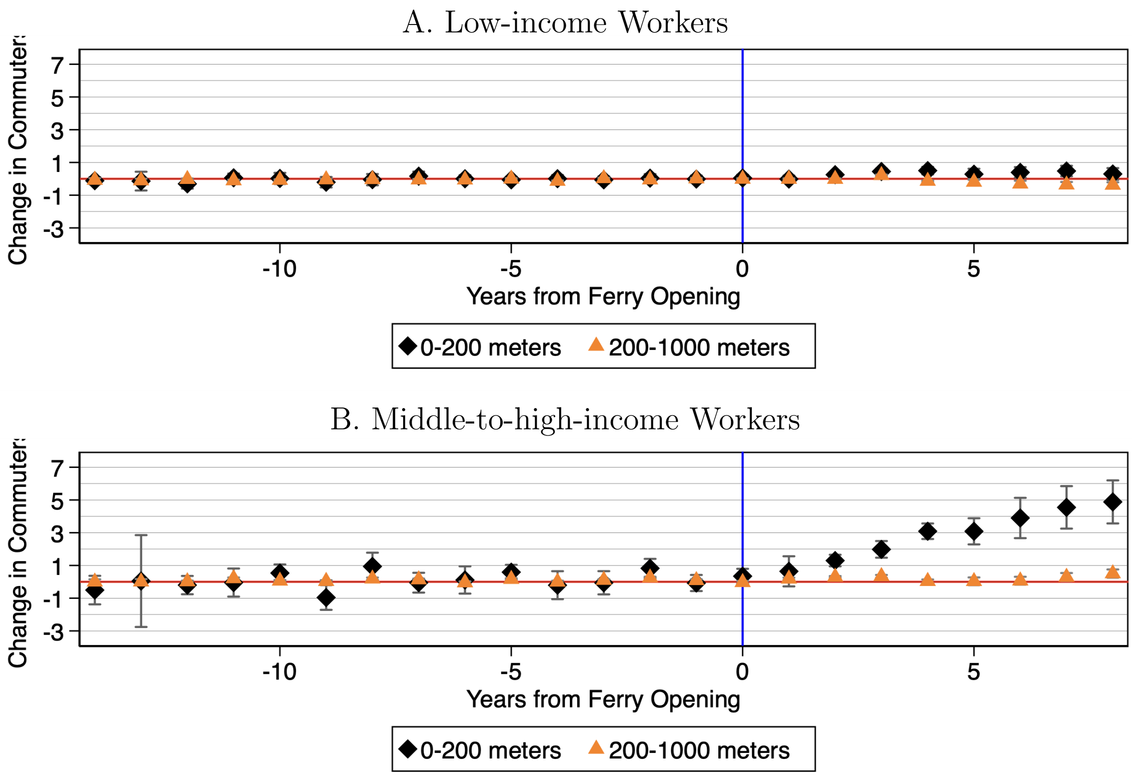


The sample of routes is divided according to whether the route was connected by a ferry at the end of the study period.

Appendix B

In the main specification of route flow change I omit tracts that are between 200 and 1,000 meters of a ferry terminal to remove spillover effects and isolate a local causal effect. In this appendix, I consider the impact of the decision by implementing an alternative setup. Figure B1 provides event study results for the 0-200 meter treatment area and separately for tracts within the adjacent 200-1,000 meter buffer. I drop tracts bisected by the Second Avenue Subway extension as this might violate the parallel trend assumption. I conduct the two estimates separately, using tracts beyond 1,000 meters as a control group. Results estimated for a 200-500 meter range (not shown) are consistent with a spatial decay in the effect.

Figure B1: Event Study of Ferry Treatment Effect on Route Level Flows, Dual Treatment Areas

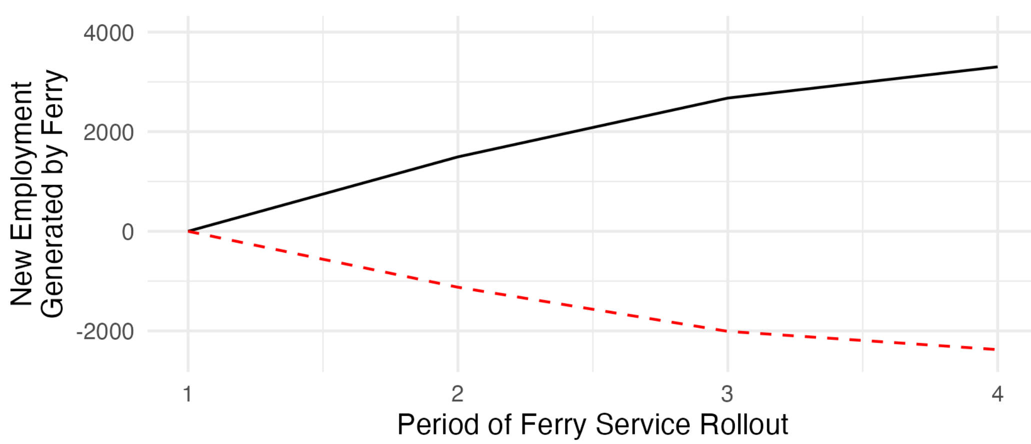


Two treatment areas are defined: tracts located 0-200 meters from an active ferry terminal, and tracts located 200-1,000 meters from an active ferry terminal. These tracts are shown in Figure 4. Estimates and confidence intervals are calculated using the methods described in Callaway and Sant'Anna (2021).

For low-income workers, while the 0-200 meter treatment effect in the final period of study was a statistically insignificant increase of 0.30 workers, the estimated impact of a ferry connection on worker flows in the 200-1,000 meter range is a marginally significant reduction in flow of 0.37 workers. For higher-income workers the 0-200 meter effect is a 4.88 worker flow increase, while tracts connected in the 200-1,000 meter treatment range saw an increase of 0.50 workers on average. Generally, the identified spillover effects are small in magnitude.

I then use the final period estimates from the Figure B1 event study to parameterize an alternative version of the full structural model, which includes buffer tracts as treated tracts. I introduce four ferry connection consumption premium parameters, rather than the two used in the main specification. The four parameters correspond to benefits for low and higher-income workers, and for the inner and outer treatment areas. Figure B2 provides the aggregate employment effect estimate from this structural analysis. When modeling spillover effects, I estimate an aggregate employment increase of 926 jobs (larger than the 700 estimated in the main analysis). I estimate more heterogeneity, with stronger positive employment effects for higher-income workers and stronger negative employment effects for low-income workers.

Figure B2: Number of Workers who Became Employed Due to Ferry Service, Dual Treatment Areas



Model results deliver estimated changes in aggregate employment. Period one corresponds to no ferry service, and subsequent periods represent each discrete expansion of ferry service, as recorded in Figure 2. This alternative model accounts for ferry effects for a 200-1,000 meter treatment range in addition to the 0-200 meter range from terminals.

Appendix C

Tables C1 and C2 reestimate the tract-level worker and job effects of a local ferry terminal using a log-linear model rather than the linear model used in Table 5 and Table 6. I find similar results.

Table C1: Effect of Ferry Connection on Local Worker Population, Log-Linear Model

	All Workers (1)	Low-income (2)	High-income (3)	F.I.R.E.† (4)	Professional Services (5)	Health Services (6)	Accom./ Food (7)
Ferry terminal dummy	0.164**	0.063	0.206**	0.241**	0.210**	0.124	0.083*
	(0.053)	(0.045)	(0.063)	(0.066)	(0.064)	(0.072)	(0.041)
Year fixed effects	Y	Y	Y	Y	Y	Y	Y
Tract fixed effects	Y	Y	Y	Y	Y	Y	Y
N	38,218	38,218	38,214	38,204	38,199	38,212	38,205
Pre-treatment average	1800.557	709.445	1091.112	298.504	233.995	243.693	121.936
% change	+17.8%	+6.5%	+22.9%	+27.3%	+23.4%	+13.2%	+8.7%

Significance levels: $*$: 5% $**$: 1%. The estimation approach follows Callaway and Sant'Anna (2021). Standard errors are shown in parentheses. †Finance, Insurance, and Real Estate.

Table C2: Effect of Ferry Connection on Local Job Counts, Log-Linear Model

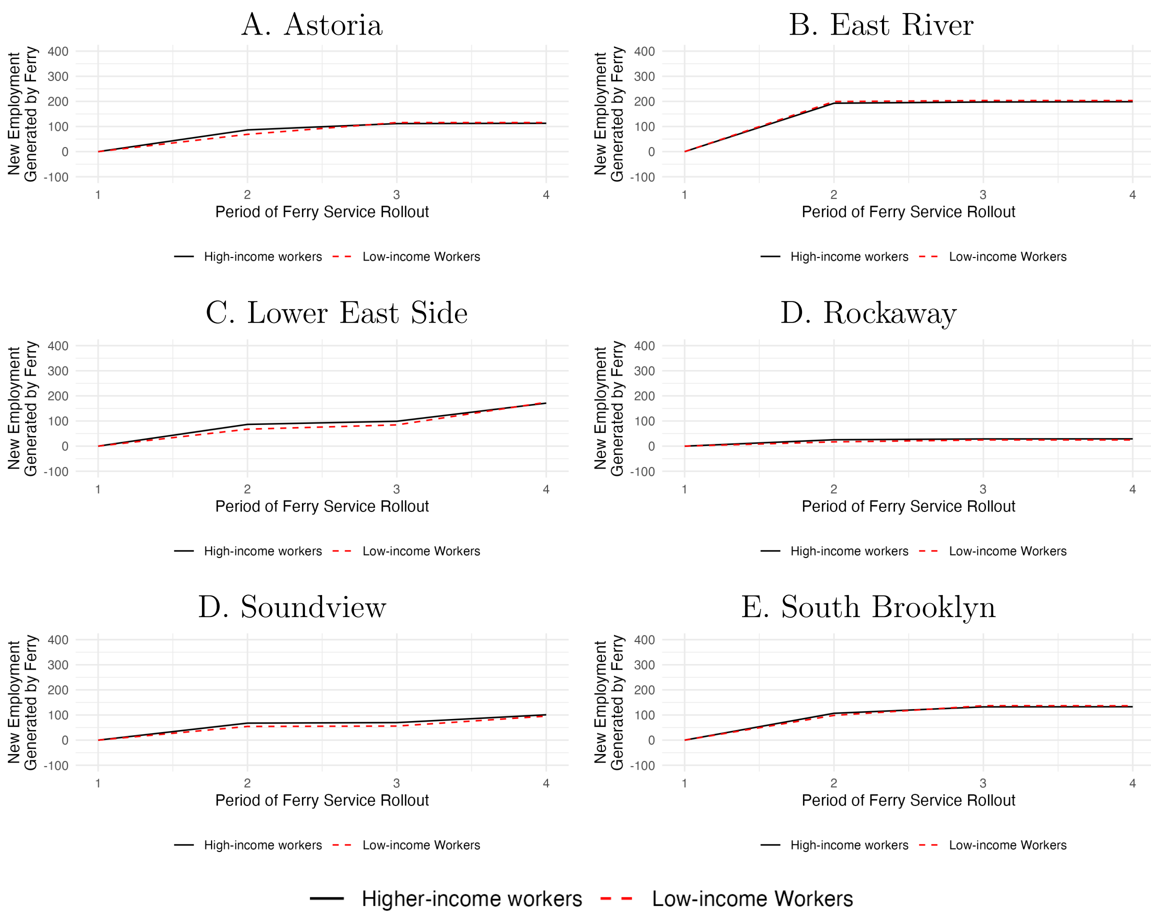
	All Workers (1)	Low-income (2)	High-income (3)	F.I.R.E.† (4)	Professional Services (5)	Health Services (6)	Accom./ Food (7)
Ferry terminal dummy	-0.019	-0.028	-0.063	0.185	0.077	-0.103	0.033
	(0.058)	(0.057)	(0.092)	(0.117)	(0.128)	(0.152)	(0.148)
Year fixed effects	Y	Y	Y	Y	Y	Y	Y
Tract fixed effects	Y	Y	Y	Y	Y	Y	Y
N	38,579	38,557	38,017	33,890	29,739	35,576	32,204
Pre-treatment average	4762.076	2183.152	2578.924	1447.363	469.487	709.297	131.248
% change	-1.9%	-2.8%	+6.5%	+20.3%	+8.0%	-10.8%	+3.4%

Significance levels: $*$: 5% $**$: 1%. The estimation approach follows Callaway and Sant'Anna (2021). Standard errors are shown in parentheses. †Finance, Insurance, and Real Estate.

Appendix D

Figure D1 provides the local employment effects of the full ferry system. Each graph captures the change in employed workers within tracts adjacent to ferry terminals that are components of each route. Ferry induced local rent increases make these areas less appealing for unemployed workers, who sort away from these areas and are replaced by employed workers. I find local employment expansion around all stations.

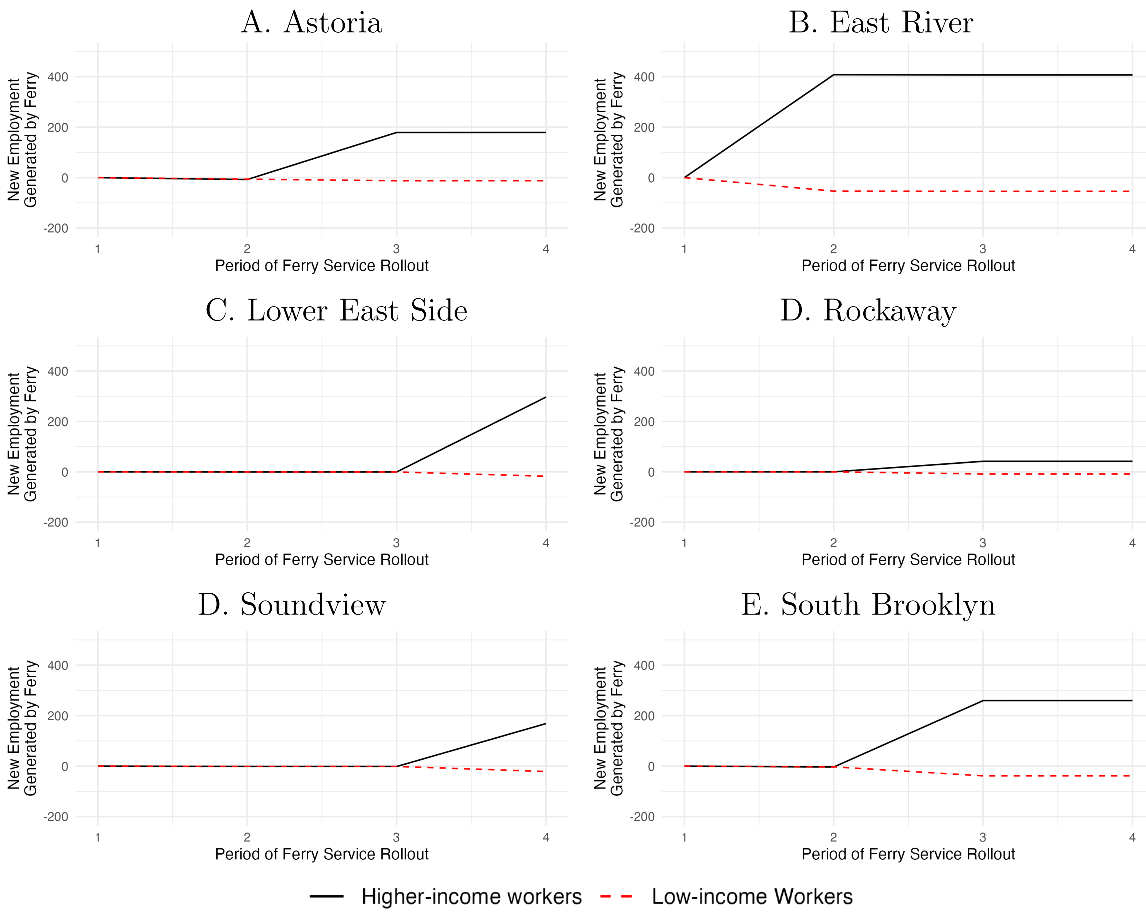
Figure D1: Number of Workers who Became Employed Due to Ferry Service, Local Impacts by Route



The graphs show the change in local employment levels within tracts treated by a ferry terminal across the phases of ferry roll out. I find all local areas near terminals experience an increase in local employment for both worker groups.

Figure D2 provides city-wide employment changes attributable to each route individually. I reestimate the model six times, but only "turn on" one ferry route at a time. All routes induce positive employment growth for higher-income workers, but marginally reduce employment among low-income workers.

Figure D2: Number of Workers who Became Employed Due to Ferry Service, Aggregate Impacts by Route

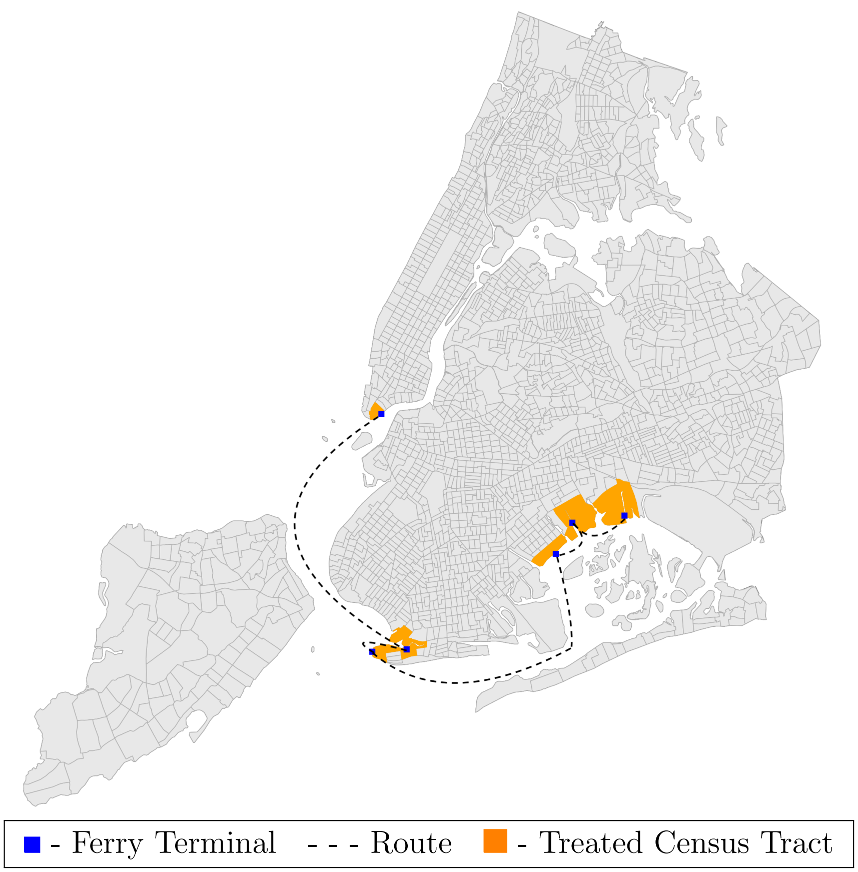


The figures show results where the model simulates only one route being active. I solve the entire model six times, but "turn on" only one route in each simulation.

Appendix E

Analysis of the NYC Ferry system shows that the system likely generated an aggregate increase in employment among higher-income workers and a small decrease among low-income workers. The pattern of diverging outcomes also held in counterfactual experiments examining each route in isolation (Figure D2). In Figure E1, I propose an imagined route that may provide stronger benefits for low-income workers. Using the route preference parameters identified in the structural model solution, I identify a set of coastal commute routes that were particularly attractive to low-income workers relative to higher-income workers. Routes from Coney Island and East New York and ending in Lower Manhattan fit this description. I plot a plausible ferry route that services these areas. I then solve the model under a scenario where the only ferry route is the route shown in Figure E1. I maintain the preference parameters recovered in the original analysis.

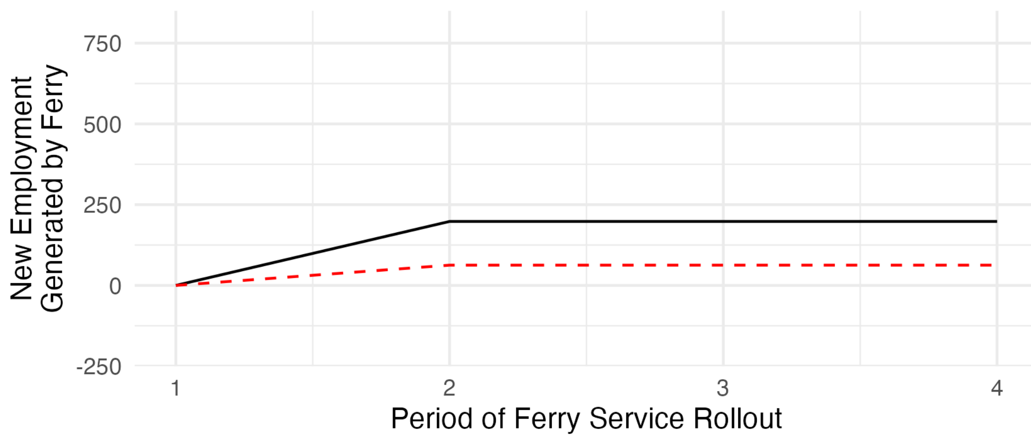
Figure E1: Census Tracts Treated by a Hypothetical Ferry Route



Tracts shown in orange are connected by an imagined new ferry route. Treated tracts are those that overlap a 200-meter buffer centered on a ferry terminal.

Figure E2 provides the city-wide employment effects under the hypothetical route. I find positive employment effects for both groups, in contrast with the estimated effects of the actual routes (Fiugre D2). The new route is estimated to raise aggregate employment by 261 jobs. Across the study period, LODES data indicates that an average of 848 low-income workers and 804 higher-income workers commuted along the routes that would be connected by this ferry. Lower uptake indicates that higher-income workers have a relatively low preference for these routes, so they are relatively less willing to move to these tracts to capture the ferry benefits. The model solution implies the ferry connections would increase rents in connected tracts by an average of \$17 per month, equivalent to a 1.4% increase.

Figure E2: Number of Workers who Became Employed Due to Ferry Service, New Hypothetical Route



Model results deliver estimated changes in aggregate employment. I imagine a new ferry route that is targeted at routes popular among low-income workers (Figure E1). The hypothetical route opens in period 2 and remains open. The model solution implies positive employment effects for both groups, with aggregate employment expanding by 198 higher-income workers and 63 low-income workers.

The results of this appendix rely on the estimated ferry connection consumption premiums (Table 8). If the preference for a ferry connection became larger for low-income workers as a result of the alternative placement of terminals, the Figure E2 results may underestimate the benefits accruing to low-income workers.