

Can Platform Size Increase Geographical Inequity? Spatial Network Externalities in Ride Sharing

Soheil Ghili and Vineet Kumar

Yale University

August 2019

Abstract

This paper uses both empirical and theoretical methods to conduct a cross-platform study in the ridesharing market. We develop a novel model-free identification strategy with limited data requirements that allows for estimation of geographical demand-supply mismatch in transportation markets. We apply our method to ride-level multi-platform data from New York City (NYC) and show smaller platforms tend to be under-supplied in less densely populated areas. We show, using high-frequency data from NYC, that the under-supply stems from high arrival times, not high prices. We argue using a theoretical model that low platform size distorts the supply of drivers towards more densely populated areas due to network effects. Motivated by this, we estimate a minimum required platform size to avoid geographical supply distortions, which informs the current policy debate in NYC around whether ridesharing platforms should be downsized. We find the minimum-required size to be approximately 3.5M rides/month for NYC, implying that downsizing Lyft and Via –but not Uber– can increase inequity.

JEL Codes: L13; R41; D62

Keywords: Ridesharing; Spatial Markets; Transportation; Geographical Inequity

1 Introduction

Ridesharing is a fast-growing and important industry offering both real-time mobility from the consumers' perspective and flexibility from the driver's perspective. It is hard to overstate the impact of ridesharing platforms that match passengers (demand) and drivers (supply) in terms of developing a nascent industry.

We draw attention to an issue in the design of ridesharing platforms: the allocation of drivers across multiple markets (locations) that the firms operate in. This is important because the location decisions of drivers has a primary effect on consumer waiting time and ride time, impacting

consumers, drivers and platform profits. Specifically, the research questions we examine include the following:

1. How should we characterize inequity in supply of ridesharing across regions? (How) can we empirically identify inequity in supply separately from heterogeneous consumer preferences across regions using observational data? Suppose a rideshare platform has too few rides in a certain geographical area. (How) can we tell whether it is because consumers in that area have lower preferences for that platform (or rideshare in general), or whether that platform is “under-supplied” in that area, due to high prices and/or wait time?
2. To the extent that some ridesharing platforms are indeed under-supplied in some areas relative to others, what is the underlying cause? Is this due to different platform strategies? Or due to driver-response to different platform characteristics, such as size?
3. Through what mechanism can a changed platform size, *all else equal*, make the supply of ridesharing geographically more balanced with demand?
4. In a given market, is there a certain size for ridesharing platforms, for which concerns about geographical supply inequity become second order? If so, how can a policymaker get an estimate of that size using publicly available data?

We use a combination of methods to provide insight into these questions, using two different datasets for empirical analysis and a theoretical model. The first two questions are addressed by empirical analysis, whereas the rest are answered using theoretical and empirical analyses.

Our concept of supply inequity is heterogeneity across regions in the proportion of potential demand (i.e., the would-be demand if price and wait time were both zero) actually served by the firms in the market. We have two goals in our empirical analysis. The first goal is to separate out and identify geographical demand heterogeneity from geographical supply inequity. Once geographical supply inequity is separated out from demand heterogeneity, the second goal is to empirically separate out two possible sources of such inequity: platform surge pricing strategy and drivers’ strategic location decisions.

To achieve our first goal of separating supply-side geographical inequity from demand-side geographical heterogeneity, we use ride-level data from the ridesharing companies Uber, Lyft and Via operating in the New York city (NYC) market. The data includes rides within and across regions (boroughs) within the larger market of NYC. We combine these data with an identification strategy based on *cross-platform relative outflow analysis*, which to our knowledge has not been proposed in the literature.

To illustrate, in July 2017, the relative-outflow for Lyft in Staten Island was 0.64. This means for every 100 Lyft rides into Staten Island from one of the other four boroughs of NYC¹ in that period,

¹The other four boroughs are Manhattan, Brooklyn, The Bronx, and Queens.

only 64 Lyft rides left Staten Island for another borough. The same number for Uber was 0.95. The relative outflows for Lyft and Uber in Manhattan were 1.24 and 1.21 respectively. We argue that under mild and testable assumptions, these comparisons between these relative outflows suggest an under-supply of Lyft drivers in Staten Island. That is price and/or wait time for a Lyft ride in Staten Island is high enough to systematically drive potential passengers towards other options. We have two main assumptions: (i) permanent migration is negligible relative to the number of ridesharing trips; (ii) the outside option (e.g. public transportation or walking) can impact flows of rides in a platform-specific or region-specific way, but not in a platform-region specific manner.

We demonstrate how to rule out the role of consumer preferences for rideshare brands (even possibly location-specific preferences). To illustrate, the relative outflow of 0.64 for Lyft in Staten Island cannot be due to low brand awareness/preference for Lyft in that region; this number, by our first assumption, means the *same population* who chose Lyft on their way into Staten Island, were on average less likely to choose Lyft on their way out. Also, the number 0.64 cannot be because it is much easier in Staten Island to use other transportation options (say ferries) than it is elsewhere; because, by our second assumption, the ease of taking ferries should also lead to a small relative outflow for Uber in Staten Island. But in reality, the relative outflow for Uber was 0.95 in Staten Island. In summary, our assumptions imply that the 0.64 number cannot be explained by heterogeneity across boroughs in either of the two major demand factors: consumer preferences, or outside transportation options. Thus, we interpret the results from the cross-platform comparison of relative outflows as a sign of Lyft's relative *under-supply* in that area: Lyft must have had prices and/or wait times in Staten Island that were unacceptably high for some customers, steering some of the potential demand towards Uber or other options.

While the above analysis indicates that supply-side factors are of primary importance in explaining inflow / outflow variation *across platforms and regions*, it does not differentiate between different supply-side factors. Our second research question focuses on separating out the impact of platform choices like price from that of location decisions made by drivers, which is our second empirical goal. Towards this goal, we obtained relatively high-frequency data including pickup time (Estimated Time of Arrival) and Price (Surge multiplier) from both Lyft and Uber across 195 locations every 30 minutes from all boroughs of NYC in May and June of 2016. We find that Lyft's surge was not systematically higher than Uber's in Staten Island, relative to other boroughs, and was on average only 0.02 higher than Uber's in Staten Island, but about 0.045 higher than Uber elsewhere.

We find that from the consumer's perspective, Lyft's pickup times (or ETA, estimated time of arrival) are much more skewed towards busier areas than Uber's. That is, both platforms have lower pickup times in Manhattan than in Staten Island but the gap is much bigger for Lyft. Lyft's pickup times are about 120 seconds more than Uber's in Staten Island and about 60 seconds less elsewhere, suggesting that high pickup times, not prices, are behind Lyft's relative under-supply

in less busy areas. But are those high pickup-times, themselves caused by prices? Or by driver behavior?

We use two approaches to rule out a primary role for price in causing these differences in pickup time. First, we use causal estimates of price on pickup time demonstrated by Cohen et al. (2016), and find that price variations can only explain $\approx 1\%$ of the variation in pickup time. Second, we examine regions / periods where all platforms have no surge pricing, and find that the above mentioned pickup time patterns still hold. Thus, we can empirically rule out the impact of price as a primary driver of pickup times.

The set of empirical analyses employed above indicate that geographical supply inequity can result from factors other than platform strategies like pricing. But what factors might cause inequity? To examine why inequity arises, we develop a theoretical model of drivers' location choices across regions in a market. It shows (i) how drivers create geographical inequity by "excessively clustering" in busier areas; and (ii) how smaller platform size, *ceteris paribus*, worsens the inequity.²

Drivers in our model choose to operate in one of $I \geq 2$ different regions that comprise the market. We model the total driver wait time based on *idle time*, when the driver is waiting for a passenger request, and the *pickup time*, where the driver goes to the passenger's pick up location. Drivers choose their region to minimize their wait time, and we focus on an *all-regions equilibrium* (i.e., an equilibrium with a strictly positive number of drivers working in each region.) We find that there is an all-regions equilibrium only when the platform has a sufficient threshold of drivers. In such an equilibrium, we find that drivers overcluster, so that the higher-demand region is supplied by even more drivers than demand would suggest. Drivers have both a positive and negative externality from being located close to each other. The negative impact is due to competition, where the platform must allocate demand arrivals across nearby drivers. However, there is a positive externality where drivers who locate closer with other drivers will be assigned passengers who are closer to them. This effect is uniquely present in spatial markets (e.g., transportation, delivery, etc.) since drivers need to pick up passengers. In short, if another driver is added to a region, it increases other drivers' *idle times* but decreases their *pickup times*. We prove two crucial results. First, we show that the equilibrium allocation of drivers is "excessively clustered" towards the area with the higher demand density. To illustrate, if 80% of the demand comes from region A and 20% from region B, in the equilibrium, *strictly more than 80%* of the drivers choose to drive in region A, and *strictly less than 20%* drive in region B. The intuition behind this result is obtained by recognizing that at proportional allocation, the idle time is the same in both regions, but the pickup time is lower in the high-demand region. This feature implies that from a driver's viewpoint the higher-demand region is more attractive due to a lower wait time. Our result implies we have

²Our model is also motivated by consistent complaints from ridesharing drivers (especially those who work for smaller platforms), who indicate that in less dense areas they find it more difficult to obtain closer rides (detailed in Appendix A).

fewer drivers in the lower-demand region, leading to geographical inequity across regions.

Moreover, we find that as the platform size is *reduced* (either by having *fewer drivers* or by *proportionally fewer drivers* and *passenger arrivals*), the overclustering *increases* and the inequity is worsened. That is, relative supply of drivers between *any* two regions i, j gets more skewed towards the higher demand region, although the relative arrival rates of passengers stays the same. Crucially, we empirically test this, both within and across platforms, using rideshare data over 1.5 years (July 2017 – December 2018). We demonstrate how inequity (measured by geographical differences in relative outflows) reduces in New York for Lyft, as the platform grows in size by over 50%, from 2.2 to 3.5 million rides per month. We also show that at any given time, the geographical distribution of relative outflows is more skewed towards Manhattan for smaller platforms than it is for larger ones. We show, using a difference-in-difference setup, that the empirical observation that low platform size leads to overclustering in busier, higher-density regions is robust to a rich set of fixed effects and functional form assumptions.

Our findings have important implications for policymakers. First, we document how to characterize the degree of inequity across regions using a relative outflow metric. Second, without recruiting a critical mass of drivers, ridesharing platforms will not be able to develop and achieve presence across regions of a market. Given such an entry barrier, regulators might consider differential support for smaller platforms. Third and perhaps most topical, we find that if policymakers use a strategy of restricting platform size (in terms of number of drivers), such an approach might increase geographical inequity. In fact, such a policy might cause smaller platforms to serve only the highest-demand areas like Manhattan, and ignore less dense areas like Staten Island. This inequity holds in the case where consumer demand is unchanged by policy, as well as the case where demand decreases similar to the reduction in supply. Importantly, we augment our theoretical recommendation with an empirical measurement procedure. We develop a simple empirical tool with mild data requirements which can help policymakers determine what platform size is “large enough” to avoid a geographical mismatch between supply and demand affecting less dense areas. The idea is to identify the platform size at which the relative outflows in different regions do not change with size. Applying our analysis to NYC, we find that around 3.5 million rides per month is the *critical inequity size*, below which there will be risk of excess clustering by ridesharing platforms. Thus, if Lyft is downsized to below its mid 2018 size, it will get excessively clustered. Via in late 2018 was at a size where it is excessively clustered. For Uber, we do not expect downsizing to run the risk of geographical supply distortions.

In answering the research questions listed above, the present paper makes several novel contributions. First, our study is unique in using a cross platform study to identify geographic equity issues in ridesharing, a topic in which there is little systematic study. Second, we provide a model-free identification strategy based on relative outflows at a region-level to separate out the supply-side (driver and platform decisions) impact on availability in a region. Having data across locations and

platforms is necessary to do this, but no panel data on drivers or passengers is required. Third, we develop a theoretical model to investigate how platform size impacts geographical inequity. In typical spatial equilibrium models, firms always prefer to face fewer competitors when there is only a negative externality. However, we identify a positive externality in ridesharing so that drivers may prefer locating in regions that are densely populated with other drivers, rather than in regions where there are few others. Such a mechanism arises from the unique features of transportation (also with taxi markets) and has not been documented, to the best of our knowledge. The inductive proof techniques we develop to obtain these results could be used beyond this paper in studying geographical demand-supply imbalances in transportation markets. Fourth, we develop a practical approach that can be used by policymakers in a city (i) to test and identify whether there is geographical inequity for any platform using ride data coming into and leaving a region; and (ii) to determine the critical minimum platform size for that city necessary to avoid geographical distortion of supply away from less dense areas.

Our research has certain limitations. First, while our focus is on identifying how drivers' location choice impacts availability and pickup time, we do not directly observe these choices. We discuss how obtaining more detailed data from platforms might improve the analysis, but also point out how such data may be of limited incremental value in studying the questions examined in this paper. Second, with our empirical analysis, we are able to identify that there is an overall impact of supply-side factors like price, availability and service; but cannot separate out consumer valuation for these different aspects and do welfare analysis. Third, our theoretical model abstracts away from competition and dynamic factors like driver learning that might impact outcomes. We also do not focus on dynamic surge pricing in this analysis because the focus has been on capturing only phenomena that are of first order importance in skewing supply towards denser areas (and we show suggestive evidence that factors we abstract from are not the most important ones). Thus, enriching the model to incorporate these effects is left to future research.

2 Related Literature

This paper builds on and contributes to different strands of the literature from the empirical and theoretical work on ridesharing markets to the literature on matching markets with spatial features.

The empirical literature on ridesharing can be roughly divided into (at least) two streams. One stream is the set of papers focusing on this market as it relates to labor economics. Chen et al. (2017) examine how much workers benefit from the schedule flexibility offered by ridesharing. Cramer and Krueger (2016) study the extent to which ridesharing, compared to the traditional taxicab system, reduces the portion of time drivers are working but not driving a passenger. Chen and Sheldon (2016) examine the reaction of labor supply to the introduction of ridesharing. Buchholz et al. (2018) estimate an optimal stopping point model to study the labor supply in the taxi-cab

industry.

The second stream of empirical papers, to which our paper belongs, are those focusing on the market design aspects of the ridesharing market. Cohen et al. (2016) use rich data from Uber to estimate the demand curve for uberX, and use it to conduct a welfare analysis providing an estimate of the extent to which Uber improves consumer welfare. Buchholz (2018) examines search-inefficiencies and how they might cause a spatial demand-supply imbalance. More recent working papers (such as Shapiro (2018); Bian (2018)) extend the setup in Buchholz (2018). Frechette et al. (2018) consider the impact of density on matching frictions and show that higher density leads to more efficient matches in the taxi industry in NYC.

Our work builds upon the insights from these papers that that there is under-supply in locations with high matching frictions (Buchholz, 2018), and that low density increases matching friction (Frechette et al., 2018). We empirically study whether there is under-supply in locations with low demand density. We also study how such under-supply relates to platform size, which has not been examined by other papers.

In comparison to the literature, two aspects of our empirical strategy are worth emphasizing. First, using the relative flows analysis, we offer a solution for separating geographical supply side inequity from geographical demand-side heterogeneity. Second, to our knowledge, ours is the first paper to empirically compare ridesharing platforms to one another, and to empirically leverage the cross-platform patterns. The literature is rich with papers that make the point that due to their superior matching technology, rideshare platforms can do a much better job than taxicabs in serving under-served areas (Buchholz, 2018; Lam and Liu, 2017). We contribute to this debate by comparing rideshare options; and we empirically argue that even once equipped with real-time matching technologies, a transportation system might have shortcomings when it comes to serving certain geographical areas. We argue that those shortcomings are due to positive externalities resulting from driver co-location.

On the theoretical side, our work is related to multiple recent papers such as Banerjee et al. (2018); Afeche et al. (2018); Castro et al. (2018); Nikzad (2018); Castillo et al. (2017); Cachon et al. (2017); Bimpikis et al. (2016); Guda and Subramanian (2019) among others. Many of these papers study the spatial mismatch between supply and demand, mostly (like this paper), assuming drivers to be strategic in choosing the locations they serve. Our paper complements this literature by providing a new microfoundation for spatial-supply demand mismatch that has not received attention. That source is differential demand density across regions. We show that areas with lower demand density get even lower supply than what would be implied by their demand levels. Relatedly, we relax the implicit invariant-to-scale assumption in those papers. To our knowledge, among the papers that study spatial demand-supply mismatches, the mismatch will not change if both demand and supply are multiplied by the same factor.

Also related to our paper are Lian and van Ryzin (2019) and Nikzad (2018). The former

considers potential economies of scale in two-sided markets and how it impacts the platform’s optimal policy regarding growth. Our argument in this paper is also one of economies of scale, in that we argue larger platforms have an easier time expanding their geographical reach. Unlike Lian and van Ryzin (2019) who focus on platform’s optimal strategy in a non-spatial way, we focus on spatial inequity and its potential policy implications. Nikzad (2018) discusses complementarity among drivers albeit with a non-spatial focus. We take a similar perspective in our spatial analysis in that a driver in our model, in spite of disliking the presence of other drivers in the same region, due to increased competition, might also benefit from their presence given that those other drivers might get asked to do an undesirable pickup which otherwise would have been assigned to her.

For the sake of parsimony and motivated by the empirical evidence, we abstract from prices in our theoretical model. Pricing in ridesharing is a large but still rapidly growing field of study, with papers including but not limited to Castro et al. (2018); Guda and Subramanian (2019); Cachon et al. (2017); Bimpikis et al. (2016); Galichon and Hsieh (2017); Castillo et al. (2017) examining various aspects of it. Despite abstracting away from prices in our theory model, we believe that both the empirical and theoretical part of our model point to interesting research directions on surge pricing. Also, there are interesting connections (not directly related to prices) between our papers and some of these pricing papers. For instance, somewhat similar in spirit to the theoretical section of our paper is Castillo et al. (2017) which introduces the “wild goose chase” problem in ridesharing markets. They capture a self-reinforcing mechanism wherein most drivers are en route to pick up passengers, making them unavailable to passengers near them, leading those passengers to be assigned to drivers who are far away, in turn sustaining the issue that drivers spend most of their time en route. Though in a spatial setting, our model shares this self-reinforcement feature in the sense that drivers leave a low-density the region due to pickups being far away. Once some drivers leave, pickups are now even farther away for those drivers who have not yet left, leading more of them to exit the region, further raising the pickup times, and so on.

Another one of such theory papers with related insights to ours is Bimpikis et al. (2016). They prove an interesting result that the most profitable pricing strategy for the platform is to try to make the demand “balanced” in the sense of making the net flow of rides with passengers to each location equal to zero. The intuition is that under balancedness, drivers are highly utilized since they do not have to travel empty to pickup locations. Our paper is related in two different ways. First, we find that empirically the rideshare data can be quite imbalanced. Second, it raises the question of whether it would be still in the platform’s best interest to price restore balance, if imbalance is caused endogenously due to the density of drivers. To illustrate, if passengers use Lyft less often to exit Staten Island than they do to enter it, because of the sparsity of Lyft’s market there, would it be profitable for the platform to lower the prices and/or increase drivers’ wages in Staten Island to obtain balance? The answer is ambiguous since there is now a tradeoff between allowing drivers to come back empty from Staten Island and look for passengers in denser areas,

or pick up somebody in Staten Island where pickups are more distant.

3 Data

We leverage two sources of data in this paper. The first data source is trip-level data which is publicly available from Taxi and Limousine Commission (TLC) of New York City. For our main relative outflow analysis, we use data on Lyft, Uber, and Via trips for July 2017 - June 2018. For each trip, we know the date and time of pickup and dropoff as well as a neighborhood indicator (again both for pickup and dropoff) partitioning the NYC proper into 256 neighborhoods. We supplement this data with publicly available data from the city to determine the borough each neighborhood is located in. The trips data will be used to conduct the relative outflows analysis which will tell us about supply side inequities across geographical areas. Table 1 provides a summary of this dataset.

Table 1: Summary of Ride counts across Platforms and Boroughs
(July 2017 - December 2018)

Platform	Bronx	Brooklyn	Manhattan	Queens	Staten Island	Total
Pickups in 1000s of rides						
Lyft	3518.66	19336.19	23956.76	10925.45	437.72	58174.79
Uber	23567.13	57993.56	96084.76	34417.17	1853.50	213916.12
Via	38.49	1123.06	15716.34	248.54	0.77	17127.20
Dropoffs in 1000s of rides						
Lyft	3601.44	19364.25	22954.42	11796.50	458.18	58174.79
Uber	24202.12	58182.93	91752.35	37897.13	1881.59	213916.12
Via	53.59	1219.89	15488.35	361.44	3.93	17127.20

The second major source of data in this paper is data from estimated pickup times and surge multipliers for all products of Uber and Lyft in NYC from late May 2015 to mid June 2016. This data was obtained from the APIs of the two platforms. In the dataset, we observe the pickup times (as estimated by the platforms) and surge multipliers every 30 minutes in 195 locations across NYC proper, for all products of Uber and Lyft (though in this paper we focus only on uberX (we term this Uber) and its Lyft-equivalent (Lyft).) The unit of observation, hence, in this dataset will be the combination of (i) date (d), (ii) time of day (t), (iii) location (ℓ), and (iv) platform (i). We use this dataset to identify whether geographical inequity is due to consumer preferences or due to other supply side factors.

4 Empirical Analysis

We first use the trip-level data to conduct a relative outflow analysis to demonstrate geographic variations in accessibility, and identify whether these arise from consumer preferences or supply-side decisions by platforms or drivers. We focus on geographical disparities in access by analyzing estimated pickup time data from Uber and Lyft across different regions of New York . We document that Lyft’s pickup time increases much more rapidly than does Uber’s as we move from Manhattan to less busy areas, in particular Staten Island. We borrow estimated results from Cohen et al. (2016) to show that price differences between Uber and Lyft is unlikely to explain the observed geographical differences in Lyft’s and Uber’s pickup times across NYC. We then argue that, even beyond prices, the unavailability of Lyft in Staten Island (relative to Uber) is not likely to be caused by competitive forces or consumer preferences. There is much anecdotal data that supports our preferred explanation, based on driver choices that in turn impact availability. We also demonstrate how variations in platform size are associated with changes in inequality across regions.

4.1 Do Ridesharing Platforms Provide Unequal “access” Across Regions?

Relative Flow Analysis: The notion of access and more specifically, inequity in access is important to consumers, platforms and policymakers. In theory, ridesharing platforms might serve consumers across different regions, but in practice, the service aspects, such as wait time or availability are important factors to ensure access, could well vary significantly.

We define inequity based on the idea that the *proportion of potential demand served* is an important measure of welfare. Thus, we can define access to ride sharing services of a given platform in a given region as follows:

$$\text{Access}_{j\ell} = \frac{\text{Demand for platform } j \text{ in a location } \ell}{\text{Demand for platform } j \text{ in a location } \ell \text{ at no surge and no wait time}} \quad (1)$$

This definition helps separate demand factors (i.e., consumer preferences in region ℓ for/against the platform, the quality of other rideshare or non-rideshare transportation options in region ℓ , etc.) from supply side factors of wait time w and price p . In a sense this definition tells us how many rides are forgone due to wait time and/or price, and importantly this requires characterizing a counterfactual outcome of “no surge pricing and no wait time”. For instance, if hypothetically, Lyft has too few rides in Staten Island because people in Staten Island are less likely to be aware of its brand, it will decrease both the numerator and the denominator of the fraction above; whereas if Lyft has too few rides because its wait time is too high, it will only decrease the numerator. Thus, comparing this measure across regions could tell us about geographical inequity.

Challenges in Empirical Operationalization: As clear as this definition may be in conceptually separating supply side factors from demand side factors, it is hard to empirically measure.

It requires estimating demand functions that are at the region level and are long-run. This is impossible with data available to regulators, and likely extremely difficult even with data available to platforms. There are at least two reasons for this. First, it is very hard to measure the market shares of the outside option in a region-based way, given it is not known how many people took public transportation, “obtained a ride” from someone, or decided not to make a trip due to wait times/prices etc.

Second, estimation of potential demand is unlikely to be adequate even with data from platforms. Papers with platform data (such as Cohen et al. (2016)) measure demand as the percentage of passengers who request a ride *conditional on using the app*. This does not capture long-run dynamics where passengers learn to not use the app in certain areas because they know rides are not available there, which leads to understating the inequity between more and less busy areas.

Operationalizing Inequity Based on Relative Outflows: In order to separate geographical heterogeneity in demand factor (consumers’ brand preference/awareness, quality of outside options) from geographical inequity in supply factors (price, wait time), we develop the method of *relative outflow analysis*. We illustrate this using a simple example. Suppose we have on a day, 100 Lyft rides arriving into Staten Island (from elsewhere), and 64 rides originating from Staten Island to other regions. For Uber, suppose that the number are 100 incoming and 95 outgoing rides. The ratio between outgoing (from a region) and incoming (into a region) rides is the relative outflow $RO_{d\ell j}$ in location ℓ on date d for platform j , where $NR_{d\ell j}^{\rightarrow}$ and $NR_{d\ell j}^{\leftarrow}$ are the outgoing and incoming number of rides respectively.³

$$RO_{d\ell j} \equiv \frac{NR_{d\ell j}^{\rightarrow}}{NR_{d\ell j}^{\leftarrow}} \quad (2)$$

Thus, in the above example, Lyft has a relative outflow of 0.64 and Uber has an outflow of 0.95. We interpret the relative outflow $RO_{d\ell j}$ from Eq. (2) as a metric of *access* to a platform j in location ℓ on date d . That is, although the *absolute number of rides* in a region can differ across platforms for the following three reasons, we argue that the *relative outflows* can only differ due to the last two reasons (i.e., supply side):

1. Consumer Preferences and/or outside options: Consumers in a region like Staten Island might prefer Uber to Lyft, leading to different ride patterns. Also, there might be other factors like word of mouth and brand awareness, which might differ across regions of the market and possibly over time as well.
2. Platform Strategy: The rideshare platforms might have different surge pricing incentives for drivers in the location, leading to a disparity in inflows and outflows.⁴.

³In practice, to avoid undefined cases, we compute $RO_{d\ell j} \approx \frac{NR_{d\ell j}^{\rightarrow} + \epsilon}{NR_{d\ell j}^{\leftarrow} + \epsilon}$ where $\epsilon = 0.001$ or some small number.

⁴There might also be a baseline price difference between the platforms. Note that the baseline prices are the same

3. Driver Location Choices: Relatively more Uber drivers choose to operate in this location than Lyft drivers. Thus, availability (as measured by pickup time) is better (lower time) for Uber than for Lyft.

The interpretation of the relative outflow metric requires two assumptions. First, we assume that migration (permanent moves) across regions in a market is negligible, especially within a short timeframe (e.g. a day). Thus, the same population of consumers who go into a region also leave the region. However, for their incoming and outgoing trips, the same consumer might choose different modes of transportation, e.g. taxi, bus, subway or even walking.⁵ Second, the value of non-ridesharing options like the bus (not observed in our ride data) can be region-specific or platform-specific but not region-platform-specific.⁶

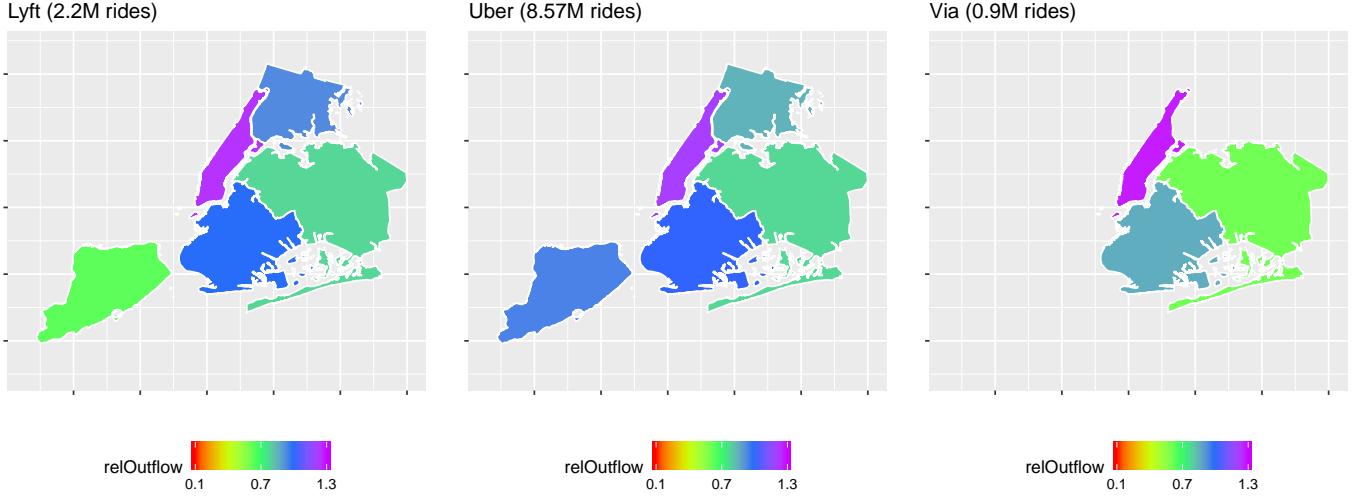
Given the above assumptions, we can rule out consumer brand preferences and outside options as contributors to relative outflow differences *across platforms* by the following logic. By our first assumption, the relative outflow of 0.64 for Lyft in Staten Island means the *same population* who chose Lyft over other options to go to Staten Island, are consistently less likely on average to make the same choice on its way out of Staten Island. This rules out consumer brand preference/awareness for the 0.64 number. It is still possible, however, that outside options are better in Staten Island relative to outside of it. This is where our second assumption is important. If people tend to not choose Lyft in Staten Island due to more convenient options such as buses, ferries, etc, then our second assumption says the same should be expected for Uber. But the relative outflow for Uber is 0.95, which implies Lyft's number must have been low due to local prices/wait times. In short, these assumptions help us identify inequity as arising due to supply-side (platform or driver) factors by ruling out heterogeneous consumer preferences as the contributing factor.

across all regions of NYC and therefore do not contribute to variation of relative outflow across regions. Moreover, in NYC as of February 2016, both platforms had virtually the same rates (Source: <https://ridesharedashboard.com/2016/02/11/comparing-uber-lyft-fare-rates-feb-2016/>)

⁵This can be due to differences between the availability of non-rideshare options across different locations. For instance, one might argue that there is a higher density of bus stops and train stations in Manhattan than elsewhere, making it easier to use those options to leave Manhattan than to enter it. One could potentially argue for the other side, positing that trains and buses may be less crowded, and, hence, more attractive, in Staten Island, making it easier to use those options to leave the Island than to enter it.

⁶It is worthwhile to understand what we are ruling out. As an illustration of this point, we are ruling out the following possibility: Lyft users who travel in and out of Staten Island tend substantially more than their Uber-user counterparts, to be from a demographic group that travels into Staten Island during hours when ferries are not available but travels out during hours when ferries are working. Once such one-directional differences among platforms are assumed away, the difference between relative outflows of Uber and Lyft in Staten Island (i.e., 0.64 v.s. 0.95) can only be attributed to relative undersupply of Lyft in Staten Island. Although we cannot directly test this assumption in our data, given that we do not see identities and/or demographics of the riders. Nevertheless, we can carry out an indirect test by comparing Uber's and Lyft's relative outflows in each hour of the day. For more on this, see Appendix C.

Figure 1: Relative Outflows for Uber, Lyft, and Via across NYC boroughs in July 2017



Does Relative Outflow vary by Platform and Region? Having described the assumptions behind the relative outflows analysis and its interpretation, we now turn to the execution of the method and interpretation of the results. Figure 1 depicts relative-outflows for each platform in July 2017 for all boroughs in which the platforms were operating at that time. As mentioned before, relative outflow measures are constructed only by looking at cross-borough rides. The relative outflow of platform X for each borough is the total number of rides from that borough to other boroughs, divided by the total number of rides from other boroughs into that borough.

Next we run a regression detailed in Eq. (3) which is consistent with the relative flow patterns shown in Figure 1.

$$\log(RO_{dli}) = \text{FE}_l + \gamma_l^{Lyft} \times \mathbf{1}_{i=\text{Lyft}} + \gamma_l^{Via} \times \mathbf{1}_{i=\text{Via}} + \varepsilon_{dli} \quad (3)$$

The results are detailed in Table 2. The only borough for which Lyft's relative outflow is significantly (at the 0.99 confidence level) and substantially different from Uber's is Staten Island. For Via, however, all three coefficients are significant and especially the negative ones (i.e., those for Brooklyn and Queens) are fairly large in magnitude. Interpreting relative outflows as measures of access, this implies that in July 2017, Lyft was as accessible as Uber in NYC in all boroughs but Staten Island; and Via's accessibility was heavily concentrated around Manhattan.

Based on the results shown in Figure 1 and Table 2, we cannot tell *why* Lyft was less accessible in Staten Island than Uber and why access to Via was concentrated around Manhattan.

To sum up, this section provides a *model-free* approach to identify geographical supply differences among regions in the presence of potential geographical demand confounds. This approach

Table 2: Relative Outflows Regression, July 2017

<i>Dependent variable:</i> log(Relative Outflow)		
	Estimate	(SE)
Lyft × The Bronx	0.052*	(0.031)
Lyft × Brooklyn	-0.025	(0.031)
Lyft × Manhattan	0.015	(0.031)
Lyft × Queens	-0.0005	(0.031)
Lyft × Staten Island	-0.358***	(0.032)
Via × Brooklyn	-0.251***	(0.031)
Via × Manhattan	0.107***	(0.031)
Via × Queens	-0.235***	(0.032)
Observations	9,522	
R ²	0.168	
Adjusted R ²	0.167	
Residual Std. Error	0.601 (df = 9509)	
F Statistic	147.397*** (df = 13; 9509)	

Note: *p<0.1; **p<0.05; ***p<0.01

has limited data requirements (e.g., no panel data needed on drivers or passengers) and can be generally used to identify geographical demand-supply mismatches in transportation markets.

Next, we delve deeper into what causes this geographical supply inequity; and argue that prices *do not play* a primary role in determining availability, as measured by pickup time.

4.2 What is the Impact of Prices?

In this section, we provide several empirical arguments to show that the difference in relative outflows is, for the most part, not caused by price differences across platforms and regions.

The first and shortest argument –but by no means the least powerful one– is that Via does not use surge pricing, unlike Uber and Lyft; however, as shown previously, Via has the most skewed relative outflows distribution among all three platforms. As we will show later in the paper, in 2018 when Via starts operating in Staten Island, the relative outflows for Via in Staten Island is much smaller than both other platforms (it was, for instance, about 0.13 in June 2018). This suggests that price is not likely to be the main driver for the documented geographical inequity.

We next focus on comparing Uber and Lyft using our data on surge multipliers and estimated arrival times from those two platforms. We do this in two ways. First, we consider the whole city and study how the prices and estimated arrival times differ between the two platforms across boroughs. We document that it is unlikely that prices directly turned away Lyft passengers in Staten Island, relative to other boroughs; and that it is unlikely that prices caused the high arrival time in Staten Island which turned some Lyft passengers away. In our second analysis, we focus our attention on Staten Island between 2am and 6am. In this time-region combination, neither Uber nor Lyft ever use surge pricing. Nevertheless, we show that within Staten Island between 2am and 6am, there is a huge gap between Uber’s and Lyft’s estimated arrival time, with Lyft’s increasing much more rapidly than Uber’s as we move towards less dense parts of the island. Thus, there is much more to the cross-platform cross-region variation in arrival times than can be explained by prices as the cause.

We next provide evidence that Lyft’s relative unavailability (or high wait time) in Staten Island is due to high wait times rather than high prices. We examine data from all of NYC and evaluate what part of the variance in pickup time is explained by price (surge) variation.

4.2.1 Pickup Time and Surge Pricing: All of NYC

Table 3 shows that the main source of Lyft’s low availability in Staten Island is coming from wait time and not prices. Lyft’s surge price factor is about 0.02 higher than Uber’s in Staten Island but about 0.044 higher than Uber’s elsewhere. Lyft’s pickup time (as estimated by the app) however, was about 114 seconds higher than Uber’s in Staten Island and about 66 seconds lower than Uber’s elsewhere. The gap between the two platforms’ pickup times further amplifies if we look at parts

of Staten Island farther away from Manhattan (not shown in the table). All of this suggests that to study the undersupply of Lyft in Staten Island, we need to focus on wait times.

Table 3: Wait times (seconds) and Surge Multipliers across Platforms and Areas (May-June 2016)

	Staten Island	All Other
Lyft Pickup Time	591.23	258.00
Uber Pickup Time	475.91	324.61
Lyft Surge	1.018	1.090
Uber Surge	1	1.046

Table 4: Pickup Times (in seconds) and Surge Multipliers by Area[†]

	Black Area	Pink Area
Lyft Pickup Time	652.80	250.40
Uber Pickup Time	456.59	326.76
Lyft Surge	1.012	1.100
Uber Surge	1	1.052

[†] Area Black or Pink based on Figure 2

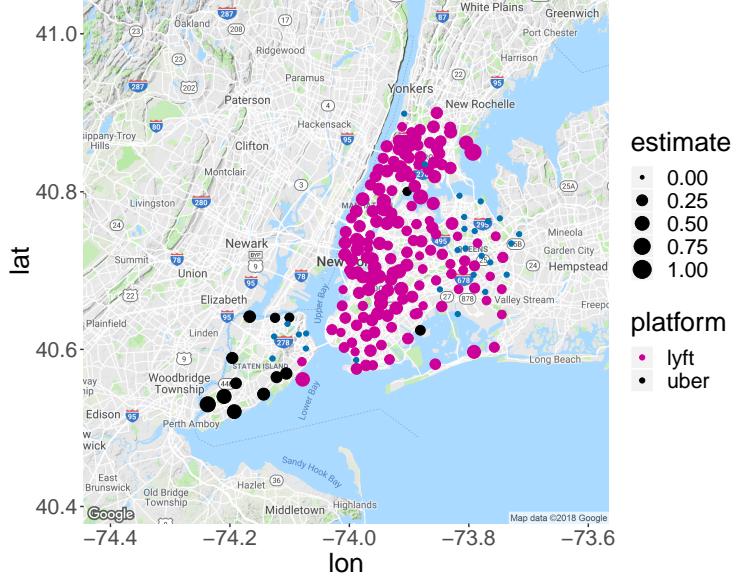
Next, we examine whether Lyft’s high wait time in Staten Island is due to its price. We first run a regression that depicts the geographical disparity between Uber’s and Lyft’s estimated pickup times across NYC, demonstrating that Lyft’s pickup time in Staten Island is substantially and statistically significantly larger than Uber’s. We then examine how much of that pickup-time disparity came from the geographical disparity in the two platforms’ prices.

The following regression characterizes the geographical disparity in pickup-time:

$$\log(t_{dt\ell j}^{\text{pickup}}) = \Delta_\ell + \delta_\ell \times \mathbf{1}_{j=\text{Lyft}} + \epsilon_{dt\ell j} \quad (4)$$

In regression (4), $t_{dt\ell j}^{\text{pickup}}$ is the estimated pickup time for platform j by its mobile app at location ℓ , time t , and date d . It is regressed on location fixed effects as well as location fixed effects interacted by a “Lyft dummy.” The interaction coefficients δ_ℓ are our coefficients of interest. If a δ_ℓ is negative, it means that at location ℓ , Lyft’s pickup time (measured in logs) is, on average, faster than Uber’s. Given that there are 195 such δ_ℓ coefficients, reporting them in a regression table might be difficult to read. Instead, we report them visually in Figure 2. Each coefficient δ_ℓ is depicted as a circle placed on a map of NYC based on the latitude and longitude of location ℓ . Wherever δ_ℓ is estimated to be negative and significant, it is colored pink in the figure. Wherever it is positive and significant, it is colored black. Statistically insignificant δ_ℓ are shown as blue dots.

Figure 2: Platform Pickup Time Regression Estimates (δ_ℓ)[†]



[†]: Lyft estimated to arrive faster in pink areas and Uber in black areas. In Blue areas, neither platform is faster than the other in a statistically significant way at the confidence level of 99.9999%.

Given that the number of coefficients is very high, we choose a very high confidence level coefficient (0.999999 instead of 0.95) to ensure our results are not due to spurious correlation.

As the regression results from Figure 2 show, areas where Uber's estimated pickup time is faster and those where Lyft is faster coincide with Staten Island and the rest of the city. Table 4 reports the average estimated pickup times for the two platforms in these areas, which given Figure 2, do not differ much from the numbers in Table 3. As detailed in this table, both Uber and Lyft have higher pickup times in the black areas (which almost coincides with Staten Island) than they do in the pink areas. However, the increases in the pickup time for Lyft as we go from the pink section to the black is much bigger than Uber's both in relative terms (a 160% increase v.s. a 40% increase) and in absolute terms (a 402 seconds increase compared to a 130 seconds increase).

We now examine the question that whether Lyft's higher pickup time in the black areas is being caused by how it is priced. More specifically, the question is whether Lyft's pricing, compared to Uber's, provides less incentive for drivers to be present in Staten Island, leading to high wait times.

To obtain causal impact of prices on pickup times, we borrow estimates on this from the literature. We adopt the *causal estimate* by Cohen et al. (2016), who use a clean Regression Discontinuity design and estimate the causal impact of a 0.1 increase in surge multiplier on estimated pickup time. They do this for uberX in four large cities (NYC, Chicago, Los Angeles, and San Francisco) in 2015. We adopt their numbers for both uberX and Lyft in NYC only, and for May/June 2016. While there is some difference in context, the fact that our results show that very little ($\approx 1\%$) of

the pickup time differences are explained by price differences, our main takeaways are robust even if the “true” causal impact of surge factor on pickup time is an order of magnitude larger than the one we borrow from Cohen et al. (2016).

The estimated causal effect by Cohen et al. (2016) is a 7.7 second decrease in pickup time for any 0.1 increase in surge multiplier. If we look again at Table 4 in light of this fact, we can see that surge differences between Uber and Lyft look too small to explain the large pickup time disparity across areas. If we were to take the causal estimate by Cohen et al. (2016) and apply it to this case, Lyft’s surge factor difference between the two areas would explain about 7 seconds out of the total 400 seconds difference between the pickup times. Also, for Uber, the near 0.05 difference between the surge factors would explain only about 4 seconds out of the total 130 seconds difference between the pickup times. Below, we formalize this argument.

Suppose the pickup time $t_{dt\ell j}^{\text{pickup}}$ is determined by the following data-generating process:

$$t_{dt\ell j}^{\text{pickup}} = a_{jdt} + b_{\ell dt} + \beta \times \text{surge}_{dt\ell j} + \nu_{dt\ell j} \quad (5)$$

This is a fairly general data-generating-process, allowing for a flexible non-parametric way by which $t_{dt\ell j}^{\text{pickup}}$ might vary as a function of jdt and/or ℓdt .

We now do three things. First, we give a formal definition for what we mean by “fraction of the pickup-time geographical disparity that is explained by prices.” Next, we provide a proposition that shows that if the data generating process for pickup times is described by equation (5), $\text{surge}_{dt\ell j}$, $t_{dt\ell j}^{\text{pickup}}$, and β values is sufficient to find this fraction. It is not necessary to know the true values of a_{jdt} or $b_{\ell dt}$. Finally, using our data on $\text{surge}_{dt\ell j}$, $t_{dt\ell j}^{\text{pickup}}$, and the value for β from Cohen et al. (2016), we show that the fraction of the geographical pickup-time differences between Uber and Lyft that is explained by prices is indeed very small.

Before the definition, we introduce two notations: $\Delta_{\text{pickup}}^{\text{DiD}}(\mathbf{a}, \mathbf{b}, \text{surge} | \beta)$ is defined to mean the difference-in-difference between Lyft and Uber pickup times across pink and black areas in Figure 2. Therefore:

$$\begin{aligned} \Delta_{\text{pickup}}^{\text{DiD}}(\mathbf{a}, \mathbf{b}, \text{surge} | \beta) &\equiv [\mathbb{E}(t_{dt\ell, \text{Lyft}}^{\text{pickup}} | l \in \mathcal{L}^{\text{Pink}}) - \mathbb{E}(t_{dt\ell, \text{Uber}}^{\text{pickup}} | l \in \mathcal{L}^{\text{Pink}})] \\ &\quad - [\mathbb{E}(t_{dt\ell, \text{Lyft}}^{\text{pickup}} | l \in \mathcal{L}^{\text{Black}}) - \mathbb{E}(t_{dt\ell, \text{Uber}}^{\text{pickup}} | l \in \mathcal{L}^{\text{Black}})] \end{aligned} \quad (6)$$

where $\mathcal{L}^{\text{Pink}}$ and $\mathcal{L}^{\text{Black}}$ are, respectively, the set of all pink and black locations in Figure 2. This number, in our data, is about 272 seconds according to Table 3.

Also define:

$$\begin{aligned} \Delta_{\text{surge}}^{\text{DiD}}(\text{surge}) &\equiv [\mathbb{E}(\text{surge}_{dt\ell, \text{Lyft}} | l \in \mathcal{L}^{\text{Pink}}) - \mathbb{E}(\text{surge}_{dt\ell, \text{Uber}} | l \in \mathcal{L}^{\text{Pink}})] \\ &\quad - [\mathbb{E}(\text{surge}_{dt\ell, \text{Lyft}} | l \in \mathcal{L}^{\text{Black}}) - \mathbb{E}(\text{surge}_{dt\ell, \text{Uber}} | l \in \mathcal{L}^{\text{Black}})] \end{aligned} \quad (7)$$

This number, in our data, is about 0.036 according to Table 3.

Definition 1. The fraction of the geographical pickup-time disparity between the two platforms, $\Delta_{\text{pickup}}^{\text{DiD}}(\mathbf{a}, \mathbf{b}, \text{surge}|\beta)$, that is causally explained by prices is defined as:

$$\Gamma \equiv \frac{\Delta_{\text{pickup}}^{\text{DiD}}(\mathbf{a}, \mathbf{b}, \text{surge}|\beta) - \Delta_{\text{pickup}}^{\text{DiD}}(\mathbf{a}, \mathbf{b}, \mathbf{1}|\beta)}{\Delta_{\text{pickup}}^{\text{DiD}}(\mathbf{a}, \mathbf{b}, \text{surge}|\beta)}$$

In words, Definition 1 says that the part of the geographical pickup-time disparity between the two platforms is the fraction by which $\Delta_{\text{pickup}}^{\text{DiD}}(\mathbf{a}, \mathbf{b}, \text{surge}|\beta)$ would change if surge factors for both platforms at all times and all places were to be set to 1.

Proposition 1. *Γ is given by: $\Gamma = \beta \times \frac{\Delta_{\text{surge}}^{\text{DiD}}(\text{surge})}{\Delta_{\text{pickup}}^{\text{DiD}}(\mathbf{a}, \mathbf{b}, \text{surge}|\beta)}$*

Proof of Proposition 1. *Follows directly from definitions.*

Proposition 1 gives us a simple way to calculate how much of the geographical difference between the two platforms in terms of their wait times can be explained by prices. The result of this calculation is:

$$\frac{7.7}{0.1} \times \frac{0.036}{272} \approx 0.0101$$

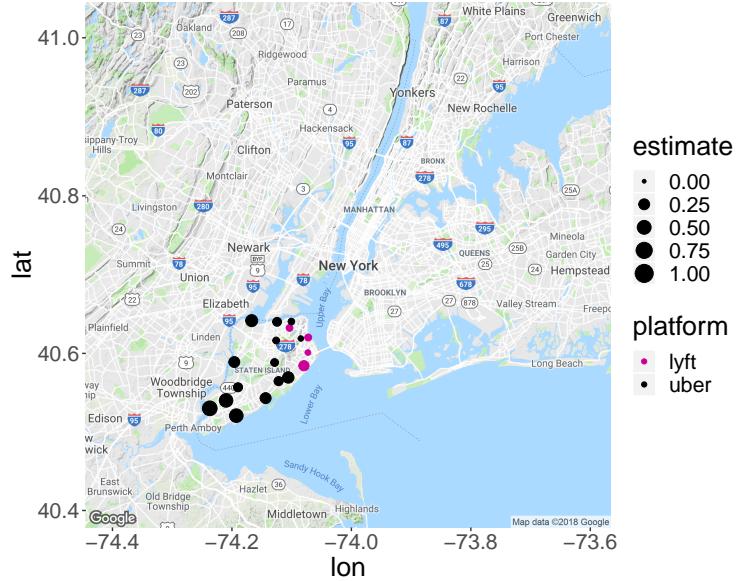
That is, if all surge multipliers were equal to 1, the geographical difference-in-differences in the platforms' pickup times would change only by 1%. Of course, this assumes that our borrowed coefficient from Cohen et al. (2016) is appropriate in our empirical context. Such an assumption may not be exactly correct. Nevertheless, the calculated value for Γ has come back so low that the majority of the pickup-time differences will still be unexplained by prices even if the true value of β is 20 times larger than 77. The above analyses suggest that price has little contribution –directly, or indirectly through impacting pickup times– to making Lyft's service less available in Staten Island. We conclude from this analysis is that *non-price factors are an important cause* of the observed differences in pickup times across platforms and regions. We therefore focus the rest of the paper on the study of non-price mechanisms.

4.2.2 Pickup Time under No Surge: Staten Island between 2am and 6am

In this section, we carry out the same analysis as we did in Section 4.2.1, except that we focus our attention to Staten Island only (as opposed to all of NYC), and 2am-6am only (as opposed to 24-hours). Neither platform does surge pricing in this location-time space. Therefore, if there is a large disparity between Uber's and Lyft's estimated arrival times across different locations within Staten Island between 2am and 6am, it cannot be caused by price differences between the two platforms. This will provide the same empirical evidence that Section 4.2.1 offers, except it does not rely on estimates from Cohen et al. (2016), which came from a slightly different context.

Figure 3 is thus restricted to 2am-6am in Staten Island (although we show all points, instead of just the statistically significant ones).

Figure 3: Platform Pickup Time Regression with no Surge in Staten Island.



†: Lyft estimated to arrive faster in pink areas and Uber in black areas.

Table 5: Pickup Times (in seconds) and Surge Multipliers by Area[‡]

	Black Area	Pink Area
Lyft Pickup Time	682.62	481.39
Uber Pickup Time	523.04	535.90
Lyft Surge	1.003	1.003
Uber Surge	1.000	1.000

‡ Area Black or Pink based on Figure 3

Table 5 summarizes the surge-factor and arrival-time comparisons across platforms and areas. It shows that while prices are constant across locations, Lyft's arrival time rapidly increases as we move towards less densely populated parts of the borough, whereas Uber's is relatively stable. As illustrated in the figure, there is almost no geographical price disparity. Uber never does surge pricing, and Lyft does it very infrequently (less than 10 instances in more than 2000 observations) and does almost equally between pink and black areas (in fact, if anything, the average surge for Lyft across black areas is by about 0.0001 higher than that in pink areas). However, the estimated arrival times are very different. Lyft's arrival time is about 54 seconds faster in pink areas, whereas Uber's is about 160 seconds faster in black areas. The difference-in-difference is about 214 seconds and is statistically significant, with a standard error of 18 seconds.

4.3 Are Uber and Lyft Geographically Segmenting the Market?

One question that might arise from our results is whether Uber and Lyft are merely differentiating their services geographically with the objective of reducing competition through spatial differentiation. Thus, one might argue that Lyft choose to focus on Manhattan, whereas Uber focuses on the suburbs. In Appendix B, we examine this possibility in detail, and demonstrate how the data patterns *do not support* this explanation.

4.4 Variation in Geographical Inequity with Platform Size

We compare two relative-outflows analyses of NYC over different timeframes, one from July 2017 and the other from June 2018 (which, respectively, are the earliest and latest months for which we have sufficient data to conduct the relative-outflows analysis.) The objective is to show the close association between small platform size and low relative outflows in less dense areas. Figure 4 demonstrates this relationship. Panel (a) shows the results from July 2017, repeating Figure 1; and panel (b) shows the results for June 2018.

Next, we translate the insight given by Figure 4 into a DiD regression, using data not only from July 2017 and June 2018, but also all other months in our dataset (from July 2017 to December 2018). The following regression allows us to determine the association between platform size on the one hand, and under-supply in lower density areas on the other hand.

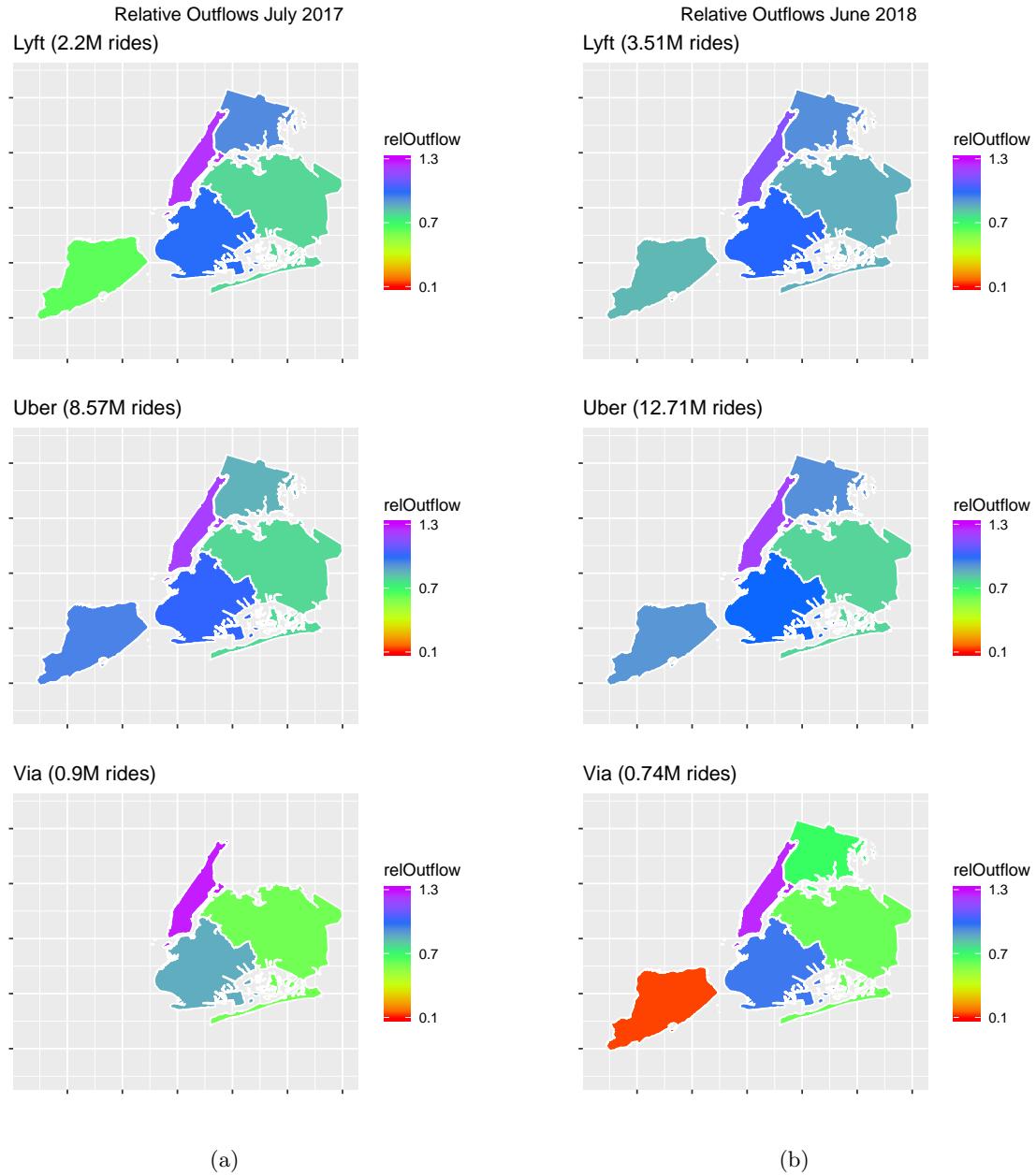
$$RO_{dlj} = \alpha_0 + \alpha_1 \log(\rho_l) + \alpha_2 \log(S_{jd}) + \alpha_3 \log(S_{jd}) \log(\rho_l) + \nu_{dlj} \quad (8)$$

where RO_{dlj} is the relative outflow for platform i at borough l on date d . Also ρ_l is the population density of borough l as of April 2019 as a proxy for demand density.⁷ Finally, S_{jd} is the size of platform j on date d , which is measured by the total number of rides given by that platform in NYC during the month in which date d occurs. Tables (6) and (7) report the results from this regression. The first table reports results when we either do not include any fixed effects in the regression or we do have fixed effects but they are not interacted. That is: platform fixed effects, year-month fixed effects, or borough fixed effects. Table 7, however, incorporates a much richer set of fixed effects. It starts, in its first three columns, fixed effects on (i) platform interacted by year-month, (ii) borough interacted by platform, and (iii) borough interacted by year-month. It then incorporates these three pairs into one single regression. Finally, the last column has interaction fixed effects among boroughs, platforms, and years (not year-month in this column).⁸

⁷We use population densities to proxy for demand densities in boroughs. We believe this assumption is reasonable, given that the rank-order of population densities across boroughs closely matches that of rides.

⁸On top of this rich set of fixed effects specifications, we also study different functional form assumptions. In equation eq. (8), we log population density and platform size but not the relative outflow. One could think of 8 different specifications here depending on which subset of these three variables are logged. We took all of these 72 regressions (8 functional form assumptions \times 9 fixed effects specifications) and the interaction coefficient is always negative with the corresponding p-value never exceeding 4×10^{-7} .

Figure 4: Relative outflows for Lyft, Uber, and Via[†]



†: Panel (a) is July 2017 and Panel (b) is June 2018

Table 6: Relative Outflow Regression with Single Fixed Effects

	<i>Dependent variable: Relative Outflow</i>			
	(1)	(2)	(3)	(4)
Constant	-7.371*** (0.142)	-7.193*** (0.212)	-7.312*** (0.142)	-6.287*** (0.118)
log(population density)	2.154*** (0.041)	2.222*** (0.040)	2.141*** (0.041)	-
log(size)	0.492*** (0.009)	0.483*** (0.014)	0.490*** (0.009)	0.448*** (0.008)
log(population density) ×				
log(size)	-0.126*** (0.003)	-0.130*** (0.003)	-0.125*** (0.003)	-0.113*** (0.002)
Fixed Effects [†]	None	P	YM	B
Observations	7,709	7,709	7,709	7,709
R ²	0.595	0.624	0.598	0.725

Note:

*p<0.1; **p<0.05; ***p<0.01

†: P:Platform, B:Borough, YM:Year-Month

Table 7: Relative Outflow regression with Interaction Fixed Effects

	<i>Dependent variable: Relative Outflow</i>				
	(1)	(2)	(3)	(4)	(5)
Constant	-7.442*** (0.166)	-9.077*** (0.496)	-6.245*** (0.129)	-19.309*** (1.310)	-6.416*** (0.833)
log(population density)	2.182*** (0.040)	2.802*** (0.147)	1.781*** (0.040)	5.963*** (0.358)	2.027*** (0.246)
log(size)	0.494*** (0.011)	0.596*** (0.036)	0.444*** (0.008)	1.326*** (0.086)	0.402*** (0.061)
log(population density) × log(size)	-0.128*** (0.003)	-0.167*** (0.011)	-0.112*** (0.002)	-0.393*** (0.023)	-0.110*** (0.018)
Fixed Effects [†]	P × YM	B × P	B × YM	(P × YM) +(P × B)+(B × YM)	P × B × Y
Observations	7,709	7,709	7,709	7,709	7,709
R ²	0.629	0.829	0.738	0.841	0.835

Note:

*p<0.1; **p<0.05; ***p<0.01

†: P:Platform, B:Borough, YM:Year-Month, Y:Year

There are two coefficients of interest for us in these tables. First, the coefficient on population density is positive and significant in all specifications. It implies that a higher relative outflow should be expected for boroughs with higher population densities.⁹ Second, and more important, is the interaction coefficient which is negative and also statistically significant across all the specifications in both tables. This coefficient indicates that as the platform gets larger, the disparity between relative outflows of low and high density boroughs shrinks. The fact that these results are robust to including platform and time fixed effects says that the association between small platform size and under-supply in low density areas can be demonstrated either (i) by looking at variation within platform over time (e.g., Lyft in July 2017 compared to Lyft in June 2018), or (ii) by looking at variation across platforms at a certain time (e.g., Lyft and Uber in July 2017).

Two observations from Figure 4 and Table 6 deserve mention here. First, in June 2018, when Lyft has reached the much larger size of about 3.5M total rides per month, the relative outflows for Uber and Lyft in Staten Island are much closer to each other than they do in July 2017 when Lyft had many fewer rides. Second, Via started operating in Staten Island and Bronx between these two time snapshots. As illustrated in panel (b) of the figure, the relative outflow for Via in Staten Island during June 2018 was 0.13, much smaller than the corresponding values for Uber and Lyft. This, together with the fact that Via is much smaller in size, suggests that platform size might be relevant in helping to reach areas with lower demand densities.

We believe that the results graphically illustrated in Figure 4 and demonstrated using regressions in Tables (6) and (7) can be interpreted as fairly causal estimates of the impact of platform size on under-supply in less dense areas, subject to the caveats we have discussed. Given that the dependent variable is the relative outflow (rather than number of pickups, etc.), factors related to geographically heterogeneous passenger preferences cannot confound the results. Also the negative coefficient on the interaction of platform size and borough population density is robustly negative and significant across a very rich set of fixed-effects specifications. The reason for such robustness is that there are multiple sources of variation all consistent with our hypothesis about the impact of platform size. Looking at cross-platform variation in a snapshot, one can see that Uber's relative outflows are more uniformly distributed across boroughs than Lyft's and Lyft's more so than Via's. Looking at within platform variation over time, once can see, for instance, that as Lyft grows, its relative outflow distribution smooths out geographically. These variations are all visible in Figure 4 with data from two sample months only; and they are shown to be significant in the regression results which uses the full data.

The robust results above can rule out a set of alternative explanations for the disparities observed among platforms in their geographical distributions of relative outflows. One such explanation is: “drivers of Lyft are less likely to live in Staten Island; hence, they are less available for pickups

⁹Obviously, this coefficient cannot be separately identified from borough fixed effects or interactions of borough fixed effects with other effects.

there.” But this can only be consistent with our results from Figure 4 and Tables (6) and (7), if in 2018, this difference between where Lyft drivers live and where Uber drivers live was inconsequential. Also, it must be that Via drivers are even more likely than both Uber and Lyft drivers to live in busier areas. Similarly, another alternative explanation for the relative under-supply of Lyft in Staten Island would be Lyft has different incentive mechanisms for their drivers in terms of where they are encouraged to drive. As we discussed before, that incentive is not likely to have been prices. In order for non-price incentives (or price incentives for that matter) to explain the under-supply of Lyft drivers as measured by relative outflows, the following reasons need to be true. It must be that the difference between the two platforms’ incentive policies dwindled over time, in a manner correlated with the movement of their size ratio, but *not because* of the movement in their size ratio.¹⁰

Another way the results from Tables (6) and (7) are useful is that it strengthens our interpretation of the cross-platform comparison among relative outflows as a supply side phenomenon. The only way the statistically significant difference between the relative outflows of Uber and Lyft can be interpreted as a demand size factor, is if there is platform-direction specific difference among passengers. That is, for instance, if Lyft passengers are more likely than Uber passengers to prefer public transportation; and in Staten Island, unlike other boroughs, public transportation is not crowded. Such hypotheses have weak support when we compare the relative outflows of the platforms across different hours of the day. Based on Figure 4 and Tables 6 and 7, even platform-direction specific differences among passengers fall short of explaining the relative outflows comparison if they are time invariant. For instance, if one believes the reason for the difference between Uber’s and Lyft’s relative outflows mid 2017 was a direction-specific difference among their passengers, then one must also believe that such direction-specific differences diminished in 2018.

5 Theoretical Model

Our theoretical model complements the empirical analysis in a number of ways. First and foremost, it describes a mechanism through which smaller platform size can cause under-supply of the ridesharing services in less busy areas. That is, the model explains why we should expect geographical supply inequity to increase as platform size decreases, *all else being fixed*. This is an important role for the theoretical analysis because although the empirical results in tables (6) and (7) suggest that all else being equal, we should expect lower size to lead to more geographical supply inequity, it is not conceptually clear why. The theory model carries out this clarification. The second role

¹⁰Also, in order for this alternative hypothesis to work, it must be that the correlation between the differential incentives provided by platforms and their size ratios is finer than yearly movements. This can be seen in column 5 of Table (7), where our coefficient of interest remains negative and significant even when we interact the platform-borough fixed effects by year dummies.

of the theory model is to produce further results that could help enrich the empirical and policy analysis. For instance, as we will show in this section, our theory model suggests that the impact of platform size on the geographical distribution of supply will satiate once the platform size becomes large enough. We will feed this insight back to the empirical analysis in the next section in order to estimate that minimum adequate size, which might be of interest to policymakers.

5.1 Setup

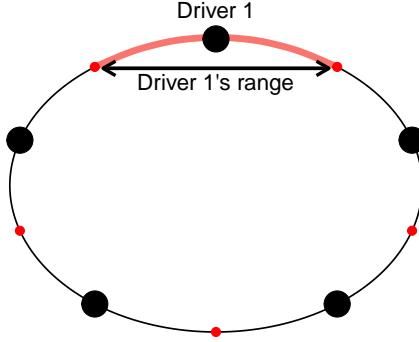
We begin by modeling a market with two regions $j = 1, 2$ with a monopolist ridesharing platform serving both (The two-regions assumption is only for the sake of intuition. Later, we will extend all our results to $I \geq 2$ regions.) The markets are circular cities a la Salop, each with consumers arriving at a rate λ_j per unit time. Each arriving passenger's location is a uniformly distributed on the circumference of the circle. There are a total N drivers who work for the platform.

Our model is a one shot game among drivers in which they simultaneously and independently choose which of the two regions to drive in. Once they choose their regions and n_j drivers pick region j , we assume for simplicity that they are identically distributed across the region (i.e., the circumference of circle j). Each driver's "catchment area" will be the arc consisting of all the points on the circle that are closer to that driver than they are to any other driver in region j . Each driver picks up the first passenger that arrives within that driver's catchment area. In practice, ridesharing platforms implement a similar matching rule.

Each driver chooses a region to drive in, minimizing her/his expected total wait time. We model two distinct reasons for the wait. First, the driver in a specific location must wait for demand realization, i.e. the arrival of a customer in the catchment area. We term this the **idle time**. Second, the driver needs to travel to the exact location of the customer to pick them up, and we call this the **pickup time**. The wait time is thus comprised of these two different components, which have divergent impact of the equilibria in ridesharing markets.

The circular city market is illustrated in Figure 5. Denote the unit travel cost by t' and the unit waiting cost as 1, to characterize driving during the pickup time to be more expensive than just waiting during the idle time. The platform allocates an arriving customer to the closest driver. When drivers are situated at equidistant points on the circumference of the city, their catchment area includes half the distance to their nearest neighbors on both sides. The expected idle time for a customer to arrive in the driver's area is $\frac{\lambda}{n}$. With a circumference of 1, the distance between drivers is $\frac{1}{n}$. Since consumer location is uniform, the distance from the driver to a consumer will be along the arc is distributed $d \sim U[0, \frac{1}{2n}]$, implying that the expected distance is $E[d] = \frac{1}{4n}$. Thus, the expected pickup time is $\frac{t'}{4n}$.

Figure 5: Circular City Model and Driver Allocation



We have the expected total wait time $W(n)$ defined from the driver's perspective as:

$$\underbrace{W_j(n)}_{\text{Total Wait Time}} = \underbrace{\frac{n}{\lambda_j}}_{\text{Idle Time}} + \underbrace{\frac{t'}{4n}}_{\text{Pickup Time}} = \left(\frac{n}{\lambda_j} + \frac{t'}{n} \right) \quad (9)$$

where $t = \frac{t'}{4}$. Observe that idle time increases in the number of drivers n since a given level of customer demand is allocated across all the drivers present in the market. The ridesharing platform allocates each customer to the closest driver. Thus, the pickup time decreases in n , since with a greater number of drivers, each driver is more likely to be allocated a passenger closer to him. This combination of idle and pickup time creates a non-monotonic U-shaped wait time function, where total wait time is initially decreasing in the number of drivers, then reaches an interior minimum, and then increases in n beyond the minimum.

Driver payoffs are characterized as $u_j = -W_j(n_j)$, so drivers will choose the market where they have lowest expected wait time. Drivers thus balance idle time and pickup time to determine which market to operate in.

Before laying out definitions of equilibria and turning to our results, we would like to reiterate that our goal in developing this theoretical analysis is to understand the mechanisms that are motivated by our empirical findings rather than in providing a comprehensive picture of the ridesharing market. With this in mind, we make a number of assumptions listed below to develop a tractable model. Most of these assumptions are made for tractability and are similar to those found in the literature. Also the proofs, especially in the case of $I > 2$ regions, are already convoluted with these assumptions, and might become intractable without them. Our assumptions are as follows:

1. Total number of drivers across both markets is fixed at N
2. Prices are the same for all markets and are high enough so that no driver finds the outside option more profitable than rideshare.
3. Drivers are undifferentiated (conditional on location) and their identity does not matter. Drivers do not have any preference for either of the regions beyond the expected wait times.

4. Consumers are located uniformly throughout the circumference.
5. The platform greedily allocates consumers to the drivers who are closest to them.
6. The allocation is modeled as continuous in n and $N - n$ across the two markets, to ensure that we can characterize equilibrium outcomes in a tractable manner.
7. There is only one platform.

Why model a monopolist firm? It may seem like assumption 7 puts the theory model at odds with the empirical analysis which studies multiple platforms. First, we note that our interest is in identifying the mechanism, and therefore we focus on the simplest possible model that can explain the empirical patterns relating platform size and inequity. We find that higher inequity can result from smaller platform size even with a monopolist platform, i.e. even in the absence of competitive effects.

What will not be feasible using such a model is *competitive* (as opposed to *comparative statics*) analysis of the market, which was not the focus of the empirical section of the paper and will not be a focus of the theory section either. We do not focus on competition since we do not believe it to be a first order determinant of why smaller platforms have more skewed supply towards busier areas. For more on this, see Appendix B.

The next subsections define market equilibria and present the results.

5.2 Defining Equilibria and Geographical Supply Inequity

We start by defining what we mean by an equilibrium of this game.

Definition 2. Under market primitives $(\lambda_1, \lambda_2, N, t)$, an allocation (n_1^*, n_2^*) of drivers between the two regions is called an equilibrium of (i) $n_1 + n_2 = N$, and (ii) no driver in each location j can strictly decrease her/his expected total wait time by choosing to drive in location $-j$.

Also, we call the allocation (n_1^*, n_2^*) a “multi region” equilibrium allocation if it is an equilibrium and if $n_1^* > 0$ and $n_2^* > 0$.

Now we define geographical supply inequity. It is equivalent to definition (1) from the empirical part of this paper, but with slight modifications to make it more suitable to the analysis in this section.

Definition 3. Under market conditions λ_1, λ_2 , we say allocation (n_1, n_2) is under-supplied in region j if we have:

$$\frac{n_j}{\lambda_j} < \frac{n_{-j}}{\lambda_{-j}}$$

The “degree of under-supply” in region j is defined by $\kappa_j = \frac{n_{-j}}{\frac{n_j}{\lambda_j}}$.

This definition simply says once you normalize the number of pickups in the two regions by their respective demand scales, if an area has a smaller number of normalized pickups, that area is under-supplied. In other words, instead of having to assume that the demand functions in the two areas are the same, it allows them to differ from one another up to a scale.

Another way to think about this definition is by comparing the ratio of drivers to the ratio of demand arrivals in the two regions. For instance, if the 80% of the demand comes from region 1 but 90% of the drivers are allocated to region 1, then region 2 is under-supplied.

5.3 Results

In this section, we will develop two important results. First, if the demand arrival rate in region 1 is strictly larger than that of region 2, then in any multi-region equilibrium, region 2 will be strictly under-supplied. Second, we show that the under-supply problem in region 2 is mitigated as the size of the platform increases, holding fixed the ratio between λ_1 and λ_2 .

First we give a lemma that helps to visually understand a multi-region equilibrium.

Proposition 2. *For a multi-region equilibrium, the wait time curves for both regions must intersect in a region where the wait time for each market is increasing in the number of drivers present in that market.*

Figure 6 helps to explain how Proposition 2 facilitates a visual grasp of conditions for multi-region equilibria. In panel (c), a multi-region equilibrium exists. This is not the case, however, in panels (a) or (b).

Next, we introduce a lemma that speaks to the existence and uniqueness of a multi-region equilibrium.

Proposition 3. *There is exactly one multi-region equilibrium if assumptions (A1) to (A3) hold. Otherwise, there is no multi-region equilibrium.*

$$(A1) \quad N \geq \sqrt{\lambda_1 t} + \sqrt{\lambda_2 t}$$

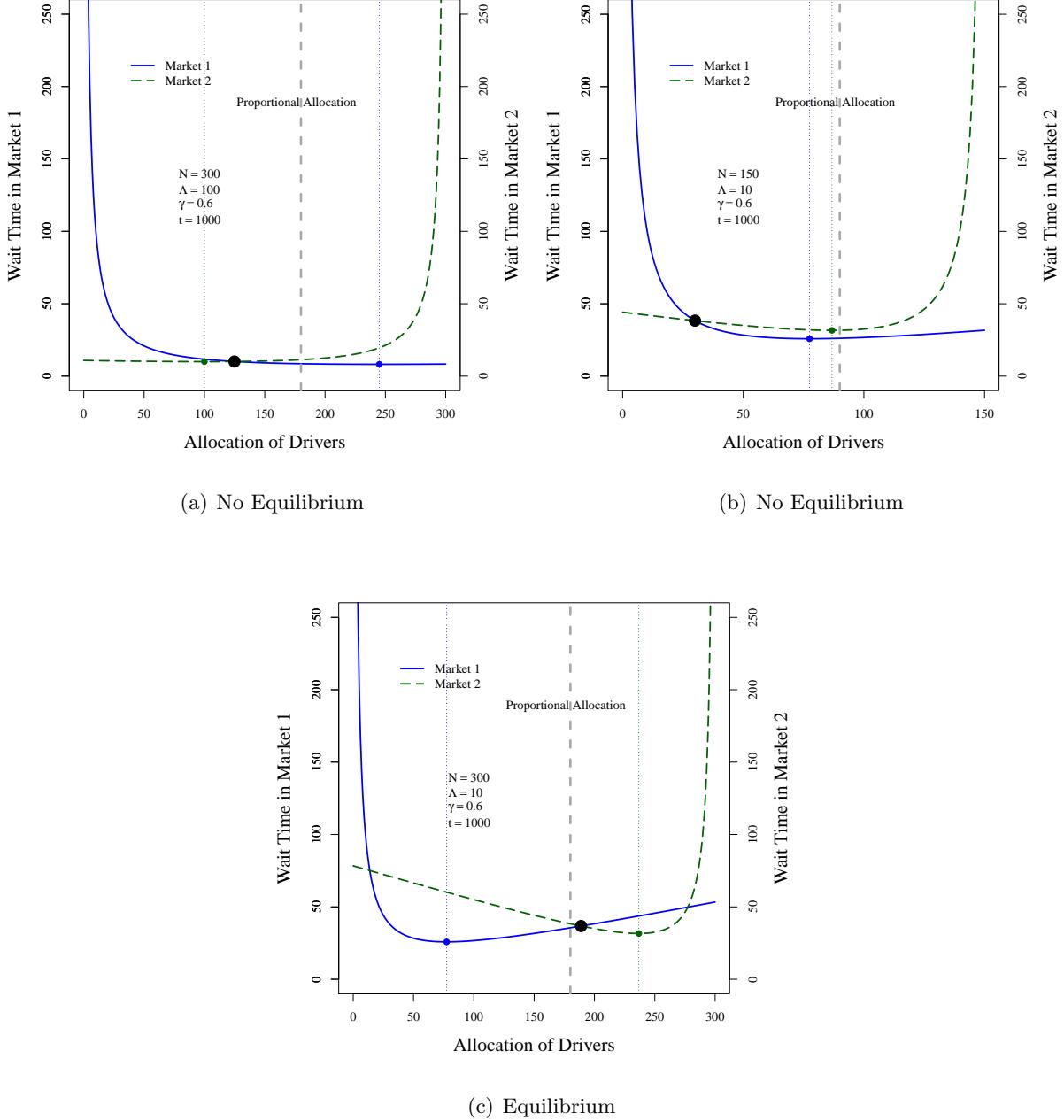
$$(A2, A3) \quad 2\sqrt{\frac{t}{\lambda_j}} \leq \frac{N - \sqrt{\lambda_j t}}{\lambda_k} + \frac{t}{N - \sqrt{\lambda_j t}} \text{ for } j = 1, 2 \text{ and } k = 3 - j$$

The first condition (A1) ensure that we have sufficient drivers to reach the minimum points of the wait time curves in both markets. The second condition (A2) requires that if market 1 is allocated the number of drivers to reach its minimum wait time $n_1 = n_1^{\min}$, and the rest of the drivers are allocated to market 2, then market 1 must have a lower wait time than market 2. The third condition (A3) is similar to (A2).

Denote the total demand by $\Lambda = \lambda_1 + \lambda_2$, and the proportion of demand present in the high-demand region 1 as $\gamma = \frac{\lambda_1}{\Lambda}$. In Figure 6, we examine three possible scenarios for the existence of

equilibria. We set parameter values for N, t, γ and Λ . In all three panels, we mark the minimum of each market's wait time by a small circle and the intersection of the curves by a large black circle. If drivers were allocated proportional to demand ($n_1 = \gamma N$), we indicate that by a vertical dashed line.

Figure 6: Wait Time and Driver Allocation



Observe that the intersection of the two wait time curves can happen in one of the following 3 regions:

1. **Both curves are decreasing** (they are below their minimum points $n_j < n_j^{min} = \sqrt{\lambda_j t}$). This corresponds to panel (a). In this case, Λ is too high relative to N , and the intersection is not an equilibrium. Assumption (A1) is not satisfied in this case.
2. **One curve is increasing whereas the other is decreasing.** Without loss of generality assume curve 1 is in decreasing region. This corresponds to panel (b). Here, N is greater than the critical threshold required and assumption **(A1)** is satisfied. However, Assumption **(A3)** is not satisfied, since at the minimum of market 2's wait time curve, the wait time in market 2 is higher than market 1. The intersection is not an equilibrium. Again, the number of drivers is insufficient to produce a multi-region equilibrium.
3. **Both curves are increasing:** They are above their minimum points ($n_j > n_j^{min} = \sqrt{\lambda_j t}$). This corresponds to panel (c). Here we have a sufficient number of drivers $N = 300$ to allocate across the two markets to produce an equilibrium with drivers in both markets. All the assumptions **(A1)**, **(A2)** and **(A3)** are satisfied. The wait time curves intersect each other at 3 points. Focusing on the middle intersection, marked by a black dot, we note that the intersection happens when both wait time curves are increasing. At this point, if a driver from market 1 moves to market 2, then she will face a higher wait time. Similarly, a driver moving from market 2 to market 1, again, will find the wait time to be higher. Thus, there is no incentive for a driver from either market to deviate to the other, and this point corresponds to the multi-region equilibrium.

We note that only in case (3) do we have a multi-region equilibrium. In the other cases, the intersection does not represent an equilibrium, as per Proposition 2.

We now turn to our main results.

Proposition 4. *Suppose that $\lambda_1 > \lambda_2$ and that a multi-region equilibrium (n_1^*, n_2^*) exists. In that case, the multi-region equilibrium is strictly under-supplied in region 2:*

$$\frac{n_1^*}{\lambda_1} > \frac{n_2^*}{\lambda_2}$$

This result coincides with our empirical observations that the relative outflow was greater in busier areas than less busy areas. This result is also consistent with the empirical observations that drivers prefer to go to areas where matching is easier (Buchholz, 2018) and that matching is easier in denser areas (Frechette et al., 2018). The combination of these two points simply implies drivers would rather gather in denser areas above and beyond what the higher demand in those areas would suggest. That is, for instance, if region 1 has 80% of the demand, then 90% of the drivers might prefer to drive in region 1.

The proof for all propositions are given in the appendix. The basic intuition is rather simple. Consider an allocation with no under-supply in either region. That is, an allocation with $\frac{n_1}{\lambda_1} = \frac{n_2}{\lambda_2}$.

Based on the expressions for idle and pickup times in those areas, it is easy to see that under such allocation, the idle times in the two regions are equal; whereas the pickup time is higher in region 2. Therefore, it would be natural to expect drivers to prefer to relocate to region 1, pushing the equilibrium in a direction in which region 2 will be under-supplied. ¹¹

We now turn to our second result which speaks to the impact of size. First, we prove that increasing the size in two different ways still obtains a multi-region equilibrium: (a) either by increasing both demand and supply or (b) by only increasing supply.

Proposition 5. *Suppose (n_1^*, n_2^*) is the multi-region EQ under $(\lambda_1, \lambda_2, N, t)$. Consider scaling up the platform size by $\phi > 1$ to $(\lambda'_1, \lambda'_2, N', t) = (\phi\lambda_1, \phi\lambda_2, \phi N, t)$. Under these new primitives, a multi-region equilibrium exists and excess clustering decreases with scaling up, i.e. $\frac{n_1'^*}{N'} < \frac{n_1^*}{N}$. In particular, as $\phi \rightarrow \infty$, under-supply in region 2 (and over-supply in region 1) tends to zero.*

This proposition speaks to the impact of platform size on geographical supply inequity in two ways. First, it shows that a scale-up in size preserves the existence of a multi-region equilibrium. This means it is possible that as size scales down, a multi-region equilibrium ceases to exist. However, as the size scales up, a multi-region equilibrium always remains in existence.

The second, more important, way our proposition talks about the impact of size is the result that increased size reduces the degree of excess clustering. To illustrate this result, it says if under $(\lambda_1, \lambda_2, N, t)$ region 1 had 80% of the total demand but n_1^* was 90% of N , under the scaled-up setting $(\lambda'_1, \lambda'_2, N', t)$ region 1 still has 80% of the total demand but will get, say, 85% of the total number of drivers. The theorem also says that if the size undergoes an extreme scale-up, then region 1 will get very close to 80% of the total number of drivers in the multi-region equilibrium.

Proposition (5) is also proved in the appendix. The intuition behind this proof is that as size gets larger and larger, the platform will get denser in both regions, reducing the importance of pickup times compared to idle times in the decision making processes of drivers. To show this, we first observe that the extent of geographical supply inequity in the equilibrium is invariant to multiplying all primitives (i.e., $\lambda_1, \lambda_2, N, t$) by the same factor γ . Therefore, the effect of multiplying only λ_1, λ_2, N by some $\gamma > 1$ on geographical supply inequity will be the same as that of dividing t by γ and holding λ_1, λ_2, N fixed. A division of t by $\gamma > 1$ means drivers care less about pickup times.

Drivers caring less about pickup times leads the equilibrium allocation to be closer to what would be implied by idle times only. It is easy to verify that if it were only the idle time that

¹¹The actual proof itself is cumbersome, given that in principle, it could be that by starting from the supply-balanced allocation and relocating more and more drivers from region 2 to region 1, we never hit a multi-region equilibrium and we finish up in a single region equilibrium with all drivers in region 1; but at the same time, a multi-region equilibrium *with under-supply in region 1* also exists. In the appendix, however, we give a proof that shows this cannot be the case, hence the multi-region equilibrium is always strictly under-supplied in the region with strictly lower demand density.

mattered to drivers, the equilibrium allocation would always be one that involved no under-supply in either region: $(n_1^*, n_2^*) = (\frac{N\lambda_1}{\lambda_1 + \lambda_2}, \frac{N\lambda_2}{\lambda_1 + \lambda_2})$. This is exactly what will be the case as the scale-up grows infinitely large.

The result in Proposition 5 is closely in line with the regression results reported in Tables (6) and (7), which based on the relative outflows analysis, show that smaller platforms have a bigger geographical supply inequity problem. It is also in line with the anecdotes discussed in Appendix A, which demonstrate that Lyft has a more difficult time serving suburbs due to long pickup times.

We add one more result regarding the impact of platform size on geographical supply inequity. Proposition 5 considers a change in a platform's size to be a scale up (or down) in both supply and demand – that is, λ_1 , λ_2 , and N . This is very helpful for connecting our theory to the empirical results in the previous sections, given that it is reasonable to assume that Lyft is smaller than Uber (and Via than Lyft) not only from the perspective of the total number of drivers, but also from the perspective of the arrival rate of potential passengers.

The above conceptualization of platform size (i.e., scale-up or down in both supply and demand), although relevant to our empirical study, may not perfectly fit an analysis of the policy debate on whether rideshare platforms need to be downsized. In that policy debate, the question is whether or not to downsize the supply side, not both the supply side and the demand side. Of course, to the extent that we believe the demand arrival rates will respond to the more limited supply, Proposition (5) does become useful in predicting that a downsize of ridesharing platforms will lead to increased under-supply in lower density areas. However, we believe it is still useful to investigate whether a similar result to Proposition (5) can be obtained if instead of scaling up both the supply side and the demand side, we scale up only the supply side. Proposition (6) below demonstrates that the answer is yes.

Proposition 6. *Suppose (n_1^*, n_2^*) is the multi-region EQ under $(\lambda_1, \lambda_2, N, t)$. If we scale up to $(\lambda_1, \lambda_2, N', t)$ for some $N' > N$, then the degree of excess clustering declines. In particular, as $N' \rightarrow \infty$, under-supply in region 2 (and in region 1) tends to zero.*

The proof for this proposition is given in the appendix. The intuition is as follows: a scale-up in N to $N' = \gamma N$ for some $\gamma > 1$ can be thought of as a combination of two changes. First, a scale-up from $(\lambda_1, \lambda_2, N, t)$ to $(\gamma\lambda_1, \gamma\lambda_2, \gamma N, t)$. Second, a scale back down in the demand arrival rates from $(\gamma\lambda_1, \gamma\lambda_2)$ to (λ_1, λ_2) . The first move is guaranteed to mitigate the geographical supply inequity problem, according to Proposition (5). The second move increases the importance of idle times (relative to pickup times) in drivers' decision making processes. Therefore, this change also shifts the new equilibrium towards what would be implied by idle times only, which would be an allocation with no geographical supply inequity.

5.4 Extension to more than two regions

The results so far describe a market in which there are two regions only. This is very helpful in understanding the basic intuition for why regions with lower demand density are relatively under-supplied. An interesting, and at the same time important, question is, however, how strongly do these result generalize to $I > 2$ regions? To illustrate, can we show for every i, j with $\lambda_i > \lambda_j$, that

- (i) $\frac{n_i^*}{\lambda_i} > \frac{n_j^*}{\lambda_j}$ and (ii) $\frac{\frac{n_i^*}{\lambda_i}}{\frac{n_j^*}{\lambda_j}}$ gets closer to 1 as N increases or as N and all λ_k proportionally increase?

Or can we prove such results only between the most or least dense areas? Or neither? Proposition (7) below delivers the positive news that the strongest generalization can be proven.

Proposition 7. Consider an I -region version (where $I > 2$) of the game set up in sections (5.1) and (5.2). The primitives are (λ, N, t) where $\lambda \in \mathbb{R}_+^I$ is the vector of demand arrival rates, N is the total number of drivers, and t is defined as before. Without loss of generality, assume that for any $i, j \in \{1, \dots, I\}$ with $i < j$, we have: $\lambda_i \geq \lambda_j$. Define an equilibrium, and an “all-region” equilibrium in the same way as previously defined for two regions. Then the following statements are true:

1. For an all-regions equilibrium, the total wait time is equal across all I regions. Also, at the equilibrium allocation, total-wait-time curves for all regions are strictly increasing in the number of drivers present in that market.
2. Any all-regions equilibrium $n^* = (n_1^*, \dots, n_I^*)$ is unique.
3. At any all-regions equilibrium, for any $i < j$, we have $\frac{n_i^*}{\lambda_i} \geq \frac{n_j^*}{\lambda_j}$. The inequality is strict if and only if $\lambda_i > \lambda_j$.
4. Suppose an all region equilibrium $n^* = (n_1^*, \dots, n_I^*)$ exists under primitives (λ, N, t) where $\lambda = (\lambda_1, \dots, \lambda_I)$. Then, if supply and demand both scale up, that is, under new primitives $(\gamma\lambda, \gamma N, t)$ with $\gamma > 1$, we have:
 - An all-regions equilibrium $n^{*\prime} = (n_1^{*\prime}, \dots, n_I^{*\prime})$ exists.
 - The new equilibrium $n^{*\prime}$ shows less geographical supply inequity than n^* in the sense that for any $i < j$, we have $\frac{\frac{n_i^{*\prime}}{\lambda_i}}{\frac{n_j^{*\prime}}{\lambda_j}} \leq \frac{\frac{n_i^*}{\lambda_i}}{\frac{n_j^*}{\lambda_j}}$. The inequality is strict if and only if $\lambda_i > \lambda_j$.
5. The same statement is true if instead of proportionally scaling up both λ and N , we scale up only N .

This proposition generalizes all of the results from section (5.3), i.e., propositions (2) through (6), to $I > 2$ regions. It does so in the strongest sense. That is, it shows that all of the results for $I = 2$ will hold pairwise once we extend to $I > 2$. This is remarkably in-line with what can be seen from the various figures and regression results in our empirical analysis. We would like

to re-emphasize the importance of these results, in particular the comparative static ones. They tell us that with growth or shrinkage of the platform, the geographical distribution of supply can spread or cluster, *without any change* in the geographical distribution of demand.

The proof of this proposition can be found in the appendix. It is based on strong induction. The basis of the induction (that is, the case of $I = 2$) is given by propositions (2) through (6). The induction works in an interrelated way. That is, for instance, in order to show item 3 from Proposition (7) holds for some $I = I_0 > 2$, we need not only assume that item 3 holds for all $I \in \{2, \dots, I_0\}$, but also that all of the other items of the proposition hold for all $I \in \{2, \dots, I_0\}$. We believe the proof techniques developed in the implementation of this induction (see appendix) can be useful beyond this paper, in the theoretical analysis of geographical demand-supply mismatch in transportation markets.¹²

6 Implications for Policy

Our work is timely since it relates to the policy debate on whether rideshare platforms should be downsized. New York and other cities have recently been considering implementing multiple policies which, either directly or indirectly, will shrink the sizes of rideshare platforms. This policy debate is important both because NYC is the largest city in the country and because of the precedent the action taken by NYC will likely set for other cities. One proposed policy is imposing a \$17 minimum hourly wage on the rideshare platforms (The Washington Post, 2018; Wired, 2019) which took effect since the beginning of February 2019 (The Hill, 2019). Another policy is to impose a cap on the number of licenses each platform can have out to drivers (hence a cap on the number of drivers who can drive for these platforms). The particular way this regulation was designed was by halting, for 12 months starting August 2018, the issuance of new licenses for drivers of rideshare platforms (The Verge, 2018; Tech Crunch, 2018). The reactions of ridesharing platforms

¹²We would also like to note, without entering the details, that the proof involves more than a straightforward application of the induction. To illustrate this, consider the case of $I = 3$. Suppose the equilibrium allocation under primitives (λ, N, t) is $n^* = (n_1^*, n_2^*, n_3^*)$. Also suppose that once we scale both N and λ up to obtain primitives $(\gamma\lambda, \gamma N, t)$, we have the equilibrium $n^{*''} = (n_1^{*''}, n_2^{*''}, n_3^{*'''})$. Assume under this new equilibrium, that $n_3^{*'''} > \gamma n_3^*$. That is, the least dense region is gaining drivers above and beyond the scale-up, as expected. This implies that regions 1 and 2 will, together, have strictly *fewer* drivers than $\gamma(n_1^* + n_2^*)$. But this renders the application of the induction to the set of regions 1 and 2 insufficient, since now those regions have undergone (i) a scale-up of γ in both demand arrival rates and total number of drivers, followed by (ii) loss of some drivers to region 3. According to our previous results, the first change reduces geographical supply inequity between regions 1 and 2; whereas the second change *increases it*. Thus, by plain application of induction, one cannot show that geographical supply side inequity between regions 1 and 2 decreases at the end. However, we prove lemmas in the appendix which guarantee the proof of the proposition, in spite of the fact that induction applies in some but not all of the cases.

to the aforementioned regulations (and potential regulations) have been mostly negative.¹³ Finally, a third approach considered by the city is to start levying a “congestion tax” on drivers. The fares for rides originating in lower Manhattan were supposed to increase by \$2.50 for taxi and \$2.75 for rideshare, effective January 1st 2019. However, the implementation has been temporarily postponed due to a lawsuit brought by a coalition of drivers and taxi owners, calling the tax “suicide charge” (The New York Times, 2019b).¹⁴ Whether this regulation will eventually be implemented is still uncertain (The New York Times, 2019a).

In this section, we discuss what we can learn from the theoretical and empirical analyses conducted in this paper for public policy issues. We focus on the potential impacts of such policies on the distribution of drivers across the city and on the geographical (in)equality of the availability of rideshare services. Of course, this by no means is a claim that geographical inequity is the only important implication of this policy. Admittedly, our paper does not focus on the labor-market consequences of this policy; nor does it focus on the impact on congestion. Nevertheless, we believe it does bring up an issue for consideration that is important in navigating future decisions.

Some policy tools might have an advantage over others from the perspective of reducing (or not increasing) geographical inequity. For instance, imposing a congestion tax (currently planned to take effect in 2020) might be preferred over downsizing of the total number of drivers. Of course, if a congestion tax leads to downsizing of rideshare platforms, it will, according to our results, also provide an incentive for drivers to drive in busier areas. However, the tax will provide a direct incentive for drivers to serve less busy areas. Such “counter-incentive” is not provided by a plain downsizing regulation. In fact, our results could be used to defend a congestion tax policy against the potential criticism that a congestion tax might cause under-supply in busier areas. Our results suggest that downsizing rideshare platforms by a congestion tax leads to driver incentives in both directions (i.e., both to drive less in busier areas, and to drive more in less busy areas), whereas a direct downsize of the number of drivers (or a geography-independent mandatory wage increase) would only increase the incentive to drive more in busy areas and less in other areas and exacerbate

¹³Uber has sued the city of New York over the year-long pause to issuing new ridesharing licenses (The Tech Crunch, 2019). Their spokesperson has claimed that such policy will do little to help mitigate the congestion in NYC (Tech Crunch, 2018). The spokesperson stated that he believed the congestion tax to be a more effective policy regarding controlling congestion. On the equity front, ridesharing platforms contend that a downsize of ridesharing will hit the outer boroughs harder than Manhattan given that those areas might have lower access to public transportation options and taxis, thereby being more reliant on ridesharing (Tech Crunch, 2018). Also, on the front of fairness among ridesharing platforms, smaller platforms have brought lawsuits against the city for multiple aspects of its crackdown on ridesharing. Lyft and Juno sued the city for the minimum wage regulation which is calculated on a weekly basis, rather than based on hours driven with a passenger. They claimed this hurts smaller platforms with lower utilization rates (Wired, 2019).

¹⁴The term comes from the fact that there have recently been multiple cases of driver suicides in NYC due to financial hardship; and that they believe some recent regulations by the city have exacerbated their situation (The New York Times, 2018)

the inequality problem.

Another qualitative takeaway from the analysis is that competition policy is complicated by driver location choice. That is, a hypothetical breakup of a large ridesharing firm into two smaller ones can have opposing effects. On the one hand, the competition between the two could benefit consumers. On the other hand, in each of those two smaller platforms (and hence, overall), the under-supply in less busy areas will increase.

On the quantitative side, we answer an interesting question motivated by our theoretical analysis. Our theory results show that geographical inequity diminishes as platform size becomes infinitely large, due to the fact that pickup times lose their importance against idle times. In a sense, this implies that if the platform size is “large enough,” then geographical inequity will not be a first order concern. A practical question is how large is this “large enough” size? To find out, we modify regression equation (8), replacing the log function applied to platform size by a function that satiates to an upper limit as the platform size increases.

We implement this by using $\log(\min(a_{Max}, \#Rides))$ instead of $\log(\#Rides)$, where a_{Max} is the parameter capturing the adequate size and is to be estimated (one could interpret a_{Max} as the size at which the impact of size on the geo-distribution of relative outflows becomes small enough so that it cannot be distinguished from noise.) We choose this way of capturing the adequate size over adopting a functional form that converges smoothly as size grows. The reason behind this choice is that we want the identification of the adequate size to come mainly from the data points at which relative outflows stop responding to platform size, as opposed to the data points at which the platform size is well below the upper limit. The regression equation implementing this notion is very similar to the earlier regression Eq. (8) on relative outflows, with the difference being the inclusion of a_{Max} . Equation (10) describes this regression:

$$RO_{d\ell j} = \alpha_0 + \alpha_1 \log(\rho_l) + \alpha_2 \log(\min(a_{Max}, S_{jd})) + \alpha_3 \log(\min(a_{Max}, S_{jd})) \log(\rho_l) + \nu_{d\ell j} \quad (10)$$

In order to make sure that the functional form of log is not substantially impacting our estimate of a_{Max} , we also estimate a version in which the size itself, as opposed to its natural log is used. Equation 11 represents this.

$$RO_{d\ell j} = \alpha_0 + \alpha_1 \log(\rho_l) + \alpha_2 \min(a_{Max}, S_{jd}) + \alpha_3 \min(a_{Max}, S_{jd}) \log(\rho_l) + \nu_{d\ell j} \quad (11)$$

Regressions (10) and (11) are estimated using non-linear least squares, and the results are reported in Table (8). The adequate size parameter, a_{Max} is estimated at 3.65M Rides/month using regression (10) and at 3.30M Rides/month using regression(11). Both estimates are statistically very significant. They are also fairly close to each other, suggesting the robustness of a_{Max} to the model specification, as we expected.

These results suggest that NYC needs to use caution if it were to downsize Lyft, and especially Via (see numbers reported in Fig. 4). Uber, on the other hand will not face distorted geographical

supply distribution if downsized. We note that given a similar dataset to what we used here, the method we laid out in this section can help identify a_{Max} in any other metropolitan area.¹⁵

Table 8: Results of regressions (10) and (11)

<i>Dependent variable: Relative Outflow</i>		
	(1)	(2)
Regression	Equation (10)	Equation (11)
α_0	-15.07*** (0.2829)	-1.449*** (0.027)
α_1	4.129*** (0.079)	6.408*** (7.604e-03)
α_2	1.030*** (0.019)	5.876e-07*** (1.254e-08)
α_3	-0.264*** (0.005)	-1.505e-07*** (3.428e-09)
a_{Max}	3.648e+06*** (7.120e+04)	3.295e+06*** (3.916e+04)
Observations	7,709	7,709

Note: *p<0.1; **p<0.05; ***p<0.01

The main coefficient of interest is a_{Max} , the adequate size for a rideshare platform to contain geographical inequity in supply.

7 Conclusion

The primary focus of this paper is to examine whether there is inequity in the allocation of drivers across regions (or availability from the consumers' viewpoint), to determine the possible sources

¹⁵We consider these estimates of a_{Max} to be lower-bounds in the sense that the minimum required size may be larger than those. The reason is, even at Uber's current size, Uber's relative outflows are skewed towards Manhattan in terms of magnitude (though less so than the other platforms.) Our empirical method does not allow to identify whether this is because of under-supply of Uber in the outer boroughs or because of geographical heterogeneity in outside options (because our method can only identify cross-platform differences.) If one believes this is due to under-supply of Uber services in the outer boroughs, then even Uber's current size would be too small, making our estimated a_{Max} a lower bound. But even then, we believe this estimate is very useful because it shows at what size the response of the geographical distribution of supply to size becomes so slow that even with an almost three-fold growth in size (from Lyft to Uber), only a negligible improvement in geographical equity of supply is achieved.

contributing to it, and to evaluate the impact of policy on the level of inequity. Towards this goal, we demonstrate how to separately identify among: (a) consumer preferences, (b) platform pricing and (c) the locational choice made by drivers. Using empirical evidence in conjunction with a novel identification strategy, we found that the location choice of drivers has more impact on inequity than either consumer preferences or pricing. We developed a theory model to understand the mechanisms that lead to inequity due to driver decisions, and found that drivers prefer high-demand locations more than underlying demand imbalances would suggest. Overclustering of drivers occurs due to positive externalities that drivers obtain from the presence of other drivers, which prevents them from obtaining rides that are too far. Finally, we identify that platform size is an important factor in determining inequity, with smaller platforms leading to higher inequity in allocation across regions.

Our research can be extended along a number of dimensions. On theory, one interesting question would be about the role of platform incentives and pricing. More specifically, it is not entirely clear whether a platform should “go along” with its drivers clustering in busier areas, or whether it should try to “correct” the clustering. On the one hand, the time spent by drivers on the way to pick a passenger up is a loss both to them and to the platform, suggesting that drivers’ action to avoid it by relocating to busy areas is in line with what the platform would want. On the other hand, a driver’s decision to relocate to another area impacts not only his/her wait times, but also other drivers’ wait times. In particular, it furthers the supply/demand mismatch, and also increases the pickup time in the quieter area more than it decreases the pickup time in the busy area. What this suggests is that the platform might want to intervene and mitigate excess clustering through prices. A theoretical model, more general than the one we built in this paper, is needed to address this question and characterize the optimal intervention by the platform.

On the empirical side, we suggest examining direct empirical evidence on drivers response to pickup times, which would require data on individual driver-level decisions to identify when, why and how much overclustering can happen. Also a similar analysis for other metropolitan areas can address whether and which rideshare platforms have reached the critical size in those markets, which would potentially benefit policymakers and firms.

References

- Afeche, P., Liu, Z., and Maglaras, C. (2018). Ride-hailing networks with strategic drivers: The impact of platform control capabilities on performance.
- Banerjee, S., Kanoria, Y., and Qian, P. (2018). The value of state dependent control in ridesharing systems. *arXiv preprint arXiv:1803.04959*.
- Bian, B. (2018). Search frictions, network effects, and spatial competition: Taxis versus uber. Technical report, Working paper, Pennsylvania State University Department of Economics.

- Bimpikis, K., Candogan, O., and Saban, D. (2016). Spatial pricing in ride-sharing networks. *Available at SSRN 2868080*.
- Buchholz, N. (2018). Spatial equilibrium, search frictions and dynamic efficiency in the taxi industry. Technical report, Working Paper.
- Buchholz, N., Shum, M., and Xu, H. (2018). Dynamic labor supply of taxicab drivers: a semiparametric optimal stopping model. *Available at SSRN 2748697*.
- Cachon, G. P., Daniels, K. M., and Lobel, R. (2017). The role of surge pricing on a service platform with self-scheduling capacity. *Manufacturing & Service Operations Management*, 19(3):368–384.
- Castillo, J. C., Knoepfle, D., and Weyl, G. (2017). Surge pricing solves the wild goose chase. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pages 241–242. ACM.
- Castro, F., Besbes, O., and Lobel, I. (2018). Surge pricing and its spatial supply response.
- Chen, M. K., Chevalier, J. A., Rossi, P. E., and Oehlsen, E. (2017). The value of flexible work: Evidence from uber drivers. Technical report, National Bureau of Economic Research.
- Chen, M. K. and Sheldon, M. (2016). Dynamic pricing in a labor market: Surge pricing and flexible work on the uber platform. In *Ec*, page 455.
- Cohen, P., Hahn, R., Hall, J., Levitt, S., and Metcalfe, R. (2016). Using big data to estimate consumer surplus: The case of uber. Technical report, National Bureau of Economic Research.
- Cramer, J. and Krueger, A. B. (2016). Disruptive change in the taxi business: The case of uber. *American Economic Review*, 106(5):177–82.
- Frechette, G. R., Lizzeri, A., and Salz, T. (2018). Frictions in a competitive, regulated market: Evidence from taxis. Technical report, National Bureau of Economic Research.
- Galichon, A. and Hsieh, Y.-W. (2017). A theory of decentralized matching markets without transfers, with an application to surge pricing.
- Guda, H. and Subramanian, U. (2019). Your uber is arriving: Managing on-demand workers through surge pricing, forecast communication, and worker incentives. *Management Science*.
- Lam, C. T. and Liu, M. (2017). Demand and consumer surplus in the on-demand economy: the case of ride sharing. *Social Science Electronic Publishing*, 17(8):376–388.
- Lian, Z. and van Ryzin, G. (2019). Optimal growth in two-sided markets. *Available at SSRN 3310559*.

Nikzad, A. (2018). Thickness and competition in ride-sharing markets. Technical report, Working paper, Stanford University.

Shapiro, M. H. (2018). Density of demand and the benefit of uber.

Tech Crunch (2018). New york city council votes to cap licenses for ride-hailing services like uber and lyft.

The Hill (2019). Uber competitors say demand has plunged under new nyc rules.

The New York Times (2018). Why are taxi drivers in new york killing themselves?

The New York Times (2019a). Congestion pricing plan for manhattan ran into politics. politics won.

The New York Times (2019b). suicide surcharge or crucial fee to fix the subway? taxi drivers brace for battle over 2.50*charge*.

The Tech Crunch (2019). Uber sues nyc to contest cap on driversrules.

The Verge (2018). In major defeat for uber and lyft, new york city votes to limit ride-hailing cars.

The Washington Post (2018). New rules guarantee minimum wage for nyc uber, lyft drivers.

Wired (2019). Lyft sues new york city over driver minimum wage rules.

Appendices

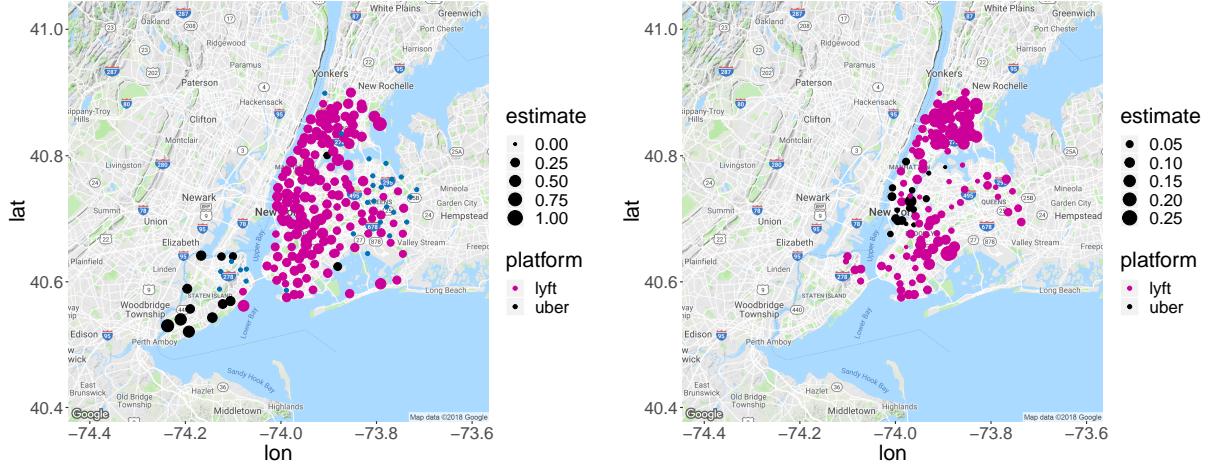
A Anecdotal Evidence from Media and Online Forums

This appendix points to a list of anecdotal pieces of evidence (by no means exhaustive) from online rideshare forums on how drivers complain about Lyft's far pickups in suburbs and how they recommend responding to it. The explanations in brackets within the quotations are from us.

- **From the online forum “Uber People,”¹⁶ a thread in the Chicago section:** The title of the thread is “To those who drive Lyft in the suburbs.” The thread was started on Dec 19 2016. The first post says “Are the ride requests you get on Lyft always seem to be far away from your location? Seems like they are always 5 miles or more for the pickup location. I got one for 12 miles last night. I drive in the Schaumburg/Palatine area [two northwestern suburbs of Chicago about 30mi away from downtown].”
- **From the same thread:** “iDrive primarily in Palatine. about two out of every five ride requests are for more than 10 minutes away. I ignore those”.
- **From the same thread:** “I was a victim of that once. Never again I take a ping more than 10 minutes away in the burbs”.
- **From the same thread:** “Yesterday was my 1st day on Lyft. Was visiting in Homer Glen [a village about 30mi southwest of downtown Chicago] & decided to try Lyft for the first time. First ping was 18 minutes away. Dang, I could make it 1/2 way downtown in that time! I ignored the ride request. 2nd ping was also 18 minutes away. Lyft app complained my acceptance rate is too low. I ignored the 2nd ping & went off-line.”
- **From the same forum, a thread titled “First 3days of Lyft”:** “If your area is spread out...and you have to take those > 10 minute requests, well...I might look for another job.”
- **From the same thread:** “Yeah, another (mostly) Lyft-specific problem, especially when working in the suburbs, is you sometimes (fairly frequently, actually) receive trip requests that are not close to your current location. I've received requests from passengers 20 miles away.”
- **From Chicago Tribune article titled “Lyft takes on Uber in suburbs”:** Jean-Paul Biondi, Chicago marketing lead for Lyft is quoted to explain the reason for Lyft's planned expansion into suburbs as follows “The main reason is we saw a lot of dropoffs in those areas, but people couldn't get picked up in those areas.” Which is in line with our reasoning that small relative-outflow is a sign of potential demand which does not get served due to under-supply.
- **From the rideshare website “Become a Rideshare Driver”:** It says successful Lyft drivers use the following strategy:
 - “The drivers usually run the Lyft app exclusively when they are in the busy downtown or city areas.”
 - “Usually in the suburbs, Uber is busier than Lyft, and in such areas, the drivers run both the Uber and Lyft apps.”

¹⁶In spite of what the name suggests, this is a general ridesharing forum, not exclusively about Uber.

Figure 7: Pickup Time and Surge for Platforms in Manhattan



B Examining “Cooperative” Geographic Segmentation

We examine whether the data provides supporting or disconforming evidence with regard to whether Uber and Lyft segment across regions to avoid competing intensely. While this is a reasonable expectation that might also serve to explain some of the regional variation in pickup times across platforms, there are a number of data patterns that run counter to this interpretation.

- **Pattern 1:** First, if we believe that the platforms are colluding across regions, shouldn't we also expect to see such patterns within a region but across sub-regions? With this goal, we examine how the pickup time for Uber and Lyft varies across the 195 sub-regions in Manhattan. If the platforms were behaving cooperatively, we might observe Lyft have a lower pickup time in one location, and Uber might have a lower pickup time in a neighboring region, for example. However, as we see in Figure 7, we see no such pattern, and note that Lyft has a lower pickup time across all sub-regions of Manhattan.
- **Pattern 2:** We can also divide Manhattan into upper and lower Manhattan, and examine the surge multipliers. We find that Uber has higher surge in lower Manhattan than Lyft, but it also has a larger pickup time. If Uber were acting in accordance to the cooperation argument made above, it would not be attempting to reduce its pickup time in lower Manhattan where Lyft has a relatively lower pickup time.
- **Pattern 3:** All of the platforms have the same relative ordering of pickup time as well as relative outflows. If the platforms were segmenting the market geographically, then we would see different relative ordering of pickup time and/or relative outflows.

Based on the above arguments, we find no support in our data for the proposition that Uber and Lyft are geographically segmenting the market.

C Hourly Analysis of Relative and Absolute Flows

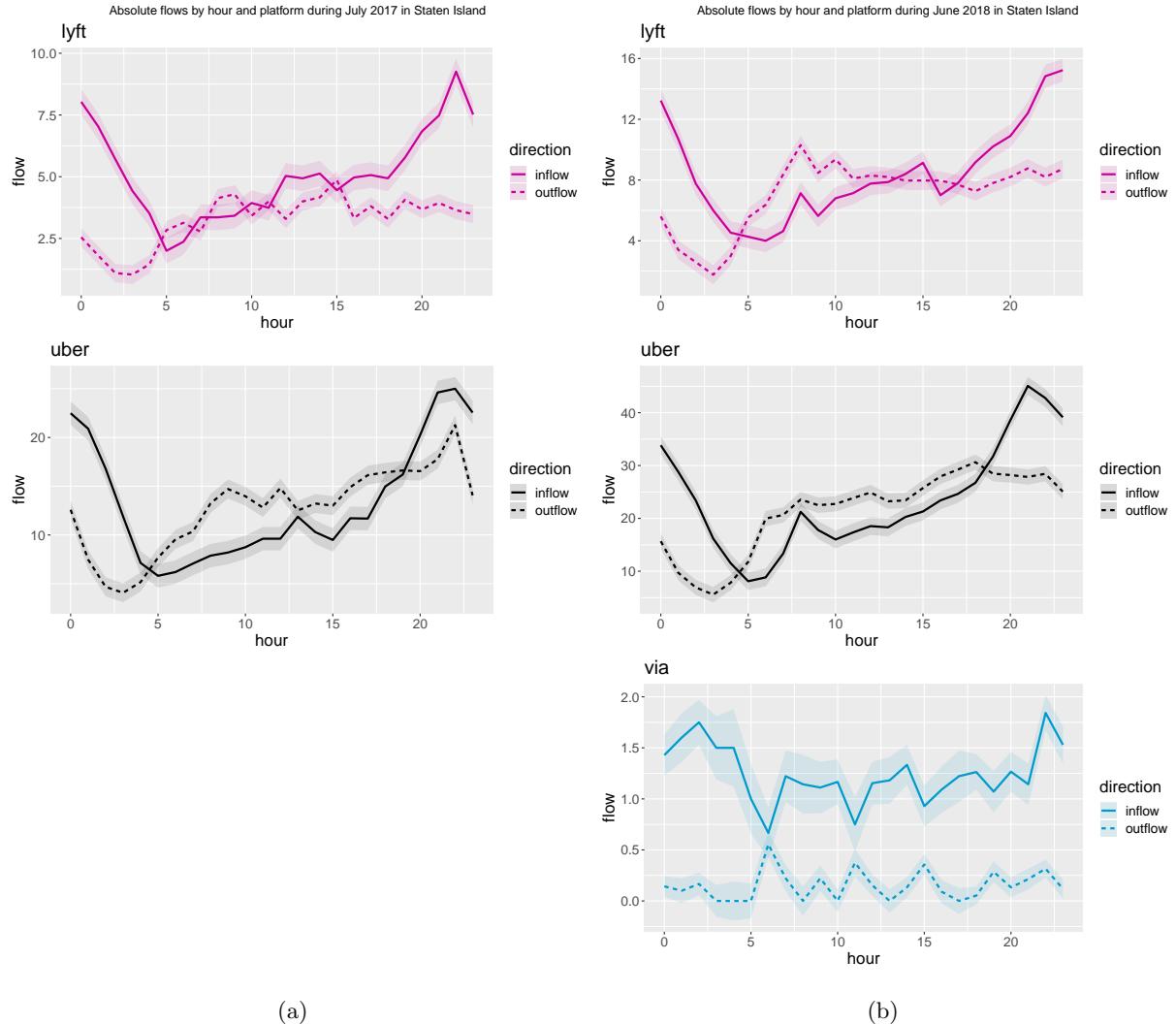
In this section, we test our identifying assumptions in the empirical part of the paper. As mentioned in Section 4, we assume that the attractiveness of the outside option to riders can change in a platform-specific or direction-specific way but *not in a platform-direction specific way*. One way this assumption can become invalid is if, say, Lyft riders are from a different demographic group than Uber riders, hence Lyft riders who enter (exit) a certain borough, do so in different hours of the day than do Uber riders. To illustrate, suppose that potential Lyft riders who want to exit Staten Island tend to need those rides during day hours when public transit options are available, whereas potential Uber riders who want to exit Staten Island tend to want to exit in early AM hours. It will, then, be natural to expect that some of the potential Lyft riders will turn to those other options, but for the potential Uber riders the outside options would be weaker. This will be a demand-side reason why the relative outflow for Lyft at Staten Island could be low, which would undermine our identification arguments.

Here, we devise a test that suggests the above concern is not a major one.¹⁷ The results compare the absolute and relative flows across the three platforms in Staten Island, averaged for each hour of the day. The comparisons are pretty consistent, suggesting that the demographic difference story outlined above cannot be behind the difference among platforms in their relative outflows (the results for other boroughs are not shown here but they also share the same consistency.) Specifically, we find that both the inflow and outflow (# of rides) as well as the ratio between them over time of day is similar for both Lyft and Uber. Thus, this data does not support the explanation that Uber and Lyft passengers have drastically different temporal preferences for ridesharing (or different schedules that leads to different availability of outside options).

Fig. 8 depicts the absolute values for inflow and outflow across time of day for the platforms.

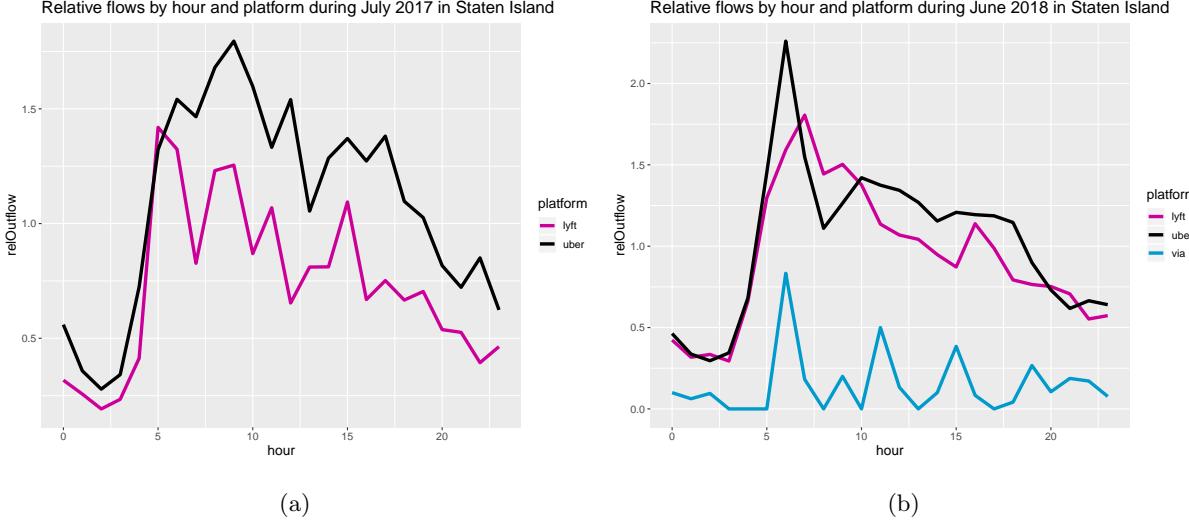
¹⁷Note that in Section 4.4 we also discussed that, from our fixed-effects regression analysis, one can argue that even if the desirability of outside options is platform-direction specific, our results will hold as long as they are not platform-direction-time specific.

Figure 8: Absolute inflows outflows for Lyft, Uber, and Via[†] in Staten Island (hourly averages)



[†]: Panel (a) is July 2017 and Panel (b) is June 2018

Figure 9: Relative outflows for Lyft, Uber, and Via[†] in Staten Island (hourly averages)



[†]: Panel (a) is July 2017 and Panel (b) is June 2018

D Proofs

Proof of Proposition 2. *If one of the curves is decreasing (say market 1 wait time is decreasing in n_1), then drivers from market 2 can reduce their wait time by switching to market 1. Note that the curves are increasing when $n_j > n_j^{\min} = \sqrt{\lambda_j t}$. ■*

Proof of Proposition 3. *First, we prove the necessary part, and then sufficiency and uniqueness.*

Necessity: *We prove necessity of (A1)-(A3) by contradiction. First, if (A1) is not satisfied, then the wait time curves can only intersect in the decreasing region for at least one market. In Figure 6, this is the case for market 1. Thus, a driver can always deviate from market 2 to market 1 and obtain a lower wait time, implying the intersection cannot be a multi-region equilibrium. Suppose condition (A2) were not true. Thus, at the minimum wait time for market 1, i.e. at the allocation $(n_1 = n_1^{\min}, n_2 = N - n_1^{\min})$, the wait time for market 1 is higher than market 2. Thus, the wait time curves can only intersect in the decreasing region for market 1, which we know cannot be a multi-region equilibrium. The necessity of (A3) is similar to (A2).*

Sufficiency: *Observe that when (A2) is true, $W_1(n_1^{\min}) < W_2(N - n_1^{\min})$. Similarly from (A3), we have $W_2(n_2^{\min}) < W_1(N - n_2^{\min})$. Also, for $n_1 > n_1^{\min}$, W_1 is an increasing function, and similar is the case for W_2 . Since we have a reversal in relative magnitude for W_1 and W_2 , we must have an intersection of the curves between n_1^{\min} and n_2^{\min} , when both wait time curves are increasing. Such an intersection permits no profitable deviation by switching to the other*

market for any driver, and is thus a multi-region equilibrium.

Uniqueness: Both wait time curves are monotonic for the region $n_1 > n_1^{\min}$ and $n_2 = (N - n_1) > n_2^{\min}$, implying that there can only be one intersection between the curves in this region.

Together, these conditions are proven equivalent to existence and uniqueness of a multi-region equilibrium. In such a case, we can characterize the multi-region equilibrium by equating the wait time distributions.¹⁸ ■

To Prove Proposition 4, we first introduce the following Lemma.

Lemma A1. When (A1)-(A3) are satisfied and when drivers are allocated proportionally to demand, the proportional allocation lies between the minimum wait times for the two regions: $n_1^{\min} < n = \gamma N < N - n_2^{\min}$.

Proof of Lemma A1. First, we prove that the proportional allocation line lies between the two minima. $n_1^{\min} = \sqrt{\lambda_1 t}$ and $n_2^{\min} = \sqrt{\lambda_2 t}$. Denote the total demand across both locations as $\Lambda = \lambda_1 + \lambda_2$ and the fraction of demand in the (higher-demand) location 1 to be $\gamma = \frac{\lambda_1}{\Lambda} > \frac{1}{2}$.

For proportional allocation to be situated between the two minimums on the graph, the following conditions need to hold:

$$\mathbf{C1: } \gamma N > n_1^{\min} = \sqrt{\lambda_1 t} = \sqrt{\gamma \Lambda t} \implies N > \sqrt{\frac{\Lambda t}{\gamma}}$$

$$\mathbf{C2: } (1 - \gamma)N > n_2^{\min} = \sqrt{\lambda_2 t} = \sqrt{(1 - \gamma)\Lambda t} \implies N > \sqrt{\frac{\Lambda t}{1 - \gamma}}$$

Observe that since $\gamma > \frac{1}{2}$, C2 \implies C1. Thus, when the demand is more skewed (higher γ), we need to have a larger platform size for condition (C2) to be satisfied.

We now prove that assumption (A1) + (A3) \implies (C2). Observe that condition that shows up is the following: Assumption (A3) implies

$$2\sqrt{\frac{t}{(1 - \gamma)\Lambda}} < \frac{N - \sqrt{(1 - \gamma)\Lambda t}}{\gamma\Lambda} + \frac{t}{N - \sqrt{(1 - \gamma)\Lambda t}}$$

The second term on the RHS can be bounded as: $\frac{t}{N - \sqrt{(1 - \gamma)\Lambda t}} < \frac{t}{\sqrt{\gamma\Lambda t}}$, since $N > \sqrt{\gamma\Lambda t} + \sqrt{(1 - \gamma)\Lambda t}$ by assumption (A1).

¹⁸In practice, we obtain the allocation equating wait times, i.e. solving $W_1(n) - W_2(N - n) = 0$, which is equivalent to identifying the roots of the polynomial equation below:

$$-n^3(\lambda_1 + \lambda_2) + n^2(2N\lambda_1 + N\lambda_2) - n(N^2\lambda_1 + 2t\lambda_1\lambda_2) + Nt\lambda_1\lambda_2 = 0$$

By Descartes' rule of signs, this equation (i.e. the numerator) has potentially 3 positive roots. In the case of multiple roots, only the one that lies between the minimum points of the wait time curves where both curves are increasing is the symmetric equilibrium. See Proposition 2.

Thus assumption (A3) implies the following:

$$\frac{N - \sqrt{(1-\gamma)\Lambda t}}{\gamma\Lambda} > 2\sqrt{\frac{t}{(1-\gamma)\Lambda}} + \frac{t}{\sqrt{\gamma\Lambda t}} \implies N > \sqrt{\Lambda t} \left(2\gamma\sqrt{\frac{1}{1-\gamma}} - \sqrt{\gamma} + \sqrt{1-\gamma} \right)$$

Next, we prove that the above inequality implies condition (C2). Observe that we need to prove the following:

$$\begin{aligned} 2\gamma\sqrt{\frac{1}{1-\gamma}} - \sqrt{\gamma} + \sqrt{1-\gamma} &> \sqrt{\frac{1}{1-\gamma}} \implies \frac{2\gamma-1}{\sqrt{1-\gamma}} - \sqrt{\gamma} + \sqrt{1-\gamma} > 0 \\ \implies \sqrt{\gamma}(\sqrt{\gamma} - \sqrt{1-\gamma}) &> 0 \end{aligned}$$

Observe that the last inequality must be true when $\gamma > \frac{1}{2}$, so (A1) + (A3) \implies (C2). ■

Proof of Proposition 4. At proportional allocation, by Lemma A1, we are in the increasing region for wait time in both markets, i.e. the proportional allocation is in between the minimum wait times for both regions. We show a symmetric equilibrium cannot be present on the left of the proportional allocation point. At proportional allocation $n = \gamma N$, market 2 wait time is higher than market 1, i.e. $W_2((1-\gamma)N) = \frac{N}{\Lambda} + \frac{t}{(1-\gamma)\Lambda} > W_1(\gamma N) = \frac{N}{\Lambda} + \frac{t}{(\gamma)\Lambda}$ since $\gamma > \frac{1}{2}$. As we move left, market 2's wait time increases further, while market 1's wait time decreases until we reach the minimum wait time for market 1, $W(n_1^{min})$. Thus, the divergence between the two markets increases. For the wait time curves to intersect, it must be in market 1's decreasing wait time region. We know from Proposition 2 that such an intersection will **not** be a multi-region equilibrium.

As we move to the right from the proportional allocation, market 1's wait time increases and market 2's wait time decreases. We know that at the minimum wait time for market 2 ($n_2 = n_2^{min}$), market 1's wait time is higher than that of market 2. By continuity, the wait time curves for the two markets must intersect in the region $\gamma N < n^* < n_2^{min}$.

In a multi-region equilibrium we equate wait times: $W_1(n_1^*) = \frac{n_1^*}{\gamma\Lambda} + \frac{t}{n_1^*}$. Since we know that $n_1^* > \gamma N$, the demand-arrival time is greater, and consequently the pickup time must be lower in market 1. This completes the proof of the proposition. ■

Lemma A2. When a multi-region equilibrium exists for a ridesharing platform with N drivers facing demand $\gamma\Lambda$ and $(1-\gamma)\Lambda$ in the two regions:

1. a multi-region equilibrium also exists when demand is unchanged and there are $N' = kN$ drivers where $k > 1$.
2. a multi-region equilibrium also exists when both the demand and number of drivers is scaled by $k > 1$ to $N' = kN$ and $\Lambda' = k\Lambda$.

Proof of Lemma A2. Consider the equivalent conditions required for the existence of a multi-region equilibrium, characterized by assumptions (A1)-(A3). Below, we show that if the conditions

are satisfied for any (N, Λ) , then they must be satisfied for (a) $(N', \Lambda') = (kN, \Lambda)$ as well as (b) $(N', \Lambda') = (kN, k\Lambda)$.

First, consider (A1). The proof of (1) is immediate. For (2), we observe that:

$$kN > \sqrt{k}\sqrt{\Lambda t} \left(\sqrt{\gamma} + \sqrt{1-\gamma} \right)$$

holds since $k > 1$ and (A1) holds.

Next, we prove (A2). The proof of (A3) is similar to that of (A2) and is omitted.

For (A2), first we denote the following function ϕ :

$$\phi(\rho) = \frac{\rho N - \sqrt{\gamma\Lambda t}}{(1-\gamma)\Lambda} + \frac{t}{\rho N - \gamma\Lambda t}$$

We prove that ϕ is increasing in ρ , or $\frac{d\phi}{d\rho} > 0$. Observe that:

$$\frac{d\phi}{d\rho} = N \left(\frac{1}{(1-\gamma)\Lambda} - \frac{t}{(\rho N - \sqrt{\gamma\Lambda t})^2} \right)$$

After some algebra and applying (A1), we obtain $\frac{d\phi}{d\rho} > 0$.

Now, for part (i), observe that setting $\rho = \frac{N'}{N} > 1$ implies that the RHS increases and the LHS does not change implying that (A2) still holds for $(N', \Lambda') = (kN, \Lambda)$.

Next, for (ii), observe that applying (A2) with $(N', \Lambda') = (kN, k\Lambda)$ gives us:

$$2\sqrt{\frac{t}{k\gamma\Lambda}} < \frac{kN - \sqrt{\gamma k\Lambda t}}{(1-\gamma)k\Lambda} + \frac{t}{kN - \gamma k\Lambda t} \text{ or } 2\sqrt{\frac{t}{\gamma\Lambda}} < \frac{\sqrt{k}N - \sqrt{\gamma\Lambda t}}{(1-\gamma)\Lambda} + \frac{t}{\sqrt{k}N - \gamma\Lambda t}$$

We need to prove the above holds whenever (A1)-(A3) hold.

Since we know that ϕ is an increasing function, we know that $\phi(\sqrt{k}) > \phi(1)$ when $k > 1$. Thus, when $(N', \Lambda') = (kN, k\Lambda)$, we find that (A2) holds for (N', Λ') .

Thus, (A1)-(A3) hold under the conditions detailed in the Lemma. ■

Proof of Proposition 5. First we note that we can reparametrize an increase in both Λ and N by $\phi > 1$ times to be equivalent to a reduction in t to $\frac{1}{\phi}t$. Then, we focus on the comparative statics with respect to t for the rest of the proof. The proof of existence of multi-region equilibrium obtains from Lemma A2 above.

The equation that characterizes a multi-region equilibrium can be parameterized in terms of $n_1 = (\alpha)N$. We know from Proposition 4 that $\alpha > \gamma$. The equilibrium condition is:

$$W^d(\alpha) \equiv W_1(\alpha N) - W_2((1-\alpha)N) = -\frac{(1-\alpha)N}{(1-\gamma)\Lambda} + \frac{(\alpha)N}{\gamma\Lambda} + \frac{t}{(1-\alpha)N} + \frac{t}{(\alpha)N} = 0 \quad (12)$$

Using the implicit function theorem, we obtain:

$$\frac{d\alpha}{dt} = -\frac{\frac{\partial W^d}{\partial t}}{\frac{\partial W^d}{\partial \alpha}} = \frac{(1-\alpha)\alpha(2\alpha-1)(1-\gamma)\gamma\Lambda}{(\alpha-1)^2\alpha^2N^2 + (2(\alpha-1)\alpha+1)(\gamma-1)\gamma\Lambda t} \quad (13)$$

The numerator is positive since $\alpha > \frac{1}{2}$. Thus, the sign of $\frac{d\alpha}{dt}$ is determined by the denominator. Below, we prove that the denominator is positive as well. The argument takes the following steps:

1. Define the denominator as $g(\alpha) = (\alpha - 1)^2 \alpha^2 N^2 + (2(\alpha - 1)\alpha + 1)(\gamma - 1)\gamma\Lambda t$.
2. Observe that $g'(\alpha) = -2(2\alpha - 1)((1 - \alpha)\alpha N^2 + (1 - \gamma)\gamma\Lambda t) < 0$, implying that $g(\alpha)$ is a decreasing function.
3. Since $\alpha \in \left[\gamma, 1 - \frac{n_2^{min}}{N}\right]$, the minimum value of $g(\alpha) = g\left(1 - \frac{n_2^{min}}{N}\right)$
4. We prove that $\min g(\alpha) = g\left(1 - \frac{n_2^{min}}{N}\right) > 0$.

$$g\left(1 - \frac{n_2^{min}}{N}\right) = \frac{\gamma^2 \Lambda t}{N^2} (N^2 - 2N\sqrt{\gamma\Lambda t} + (2\gamma - 1)\Lambda t) \quad (14)$$

$$= \frac{\gamma^2 \Lambda t}{N^2} ((N - \sqrt{\gamma\Lambda t})^2 - t\Lambda(1 - \gamma)) \quad (15)$$

where the term in parentheses is positive directly from assumption **(A1)**.

Thus, we know that $\frac{d\alpha}{dt} > 0$ implying that as t increases, the proportion of supply going to the higher-demand market is greater. ■

Proof of Proposition 6. As mentioned in the text of the paper, a scale-up in N can be thought of as a scale-up in $(\lambda_1, \lambda_2, N)$, followed by a scale back down in (λ_1, λ_2) . From Proposition 5, we know that the first scale-up (i) preserves the existence of a multi-region equilibrium and also (ii) makes it strictly less under-supplied in region 2. Therefore, the proof of Proposition (6) will be complete if we show that the second scale back down also preserves the existence of a multi-region equilibrium and makes it less under-supplied in region 2.

To see this, suppose (n_1^*, n_2^*) is the multi-region equilibrium under $(\lambda_1, \lambda_2, N, t)$. Let $\lambda'_i = \frac{\lambda_i}{\gamma}$ for $i \in \{1, 2\}$ and some $\gamma > 1$. We will now show that under $(\lambda'_1, \lambda'_2, N, t)$, there is a multi-region equilibrium with strictly less under-supply in region 2 than what is implied by (n_1^*, n_2^*) .

Lemma A3. The following statements are true about the “old” equilibrium allocation (n_1^*, n_2^*) under the “new” parameters $(\lambda'_1, \lambda'_2, N, t)$:

1. The total wait function $W_2(n)$ is strictly increasing at $n = n_2^*$.
2. At the allocation (n_1^*, n_2^*) , the wait time in region 1 is strictly higher than that in region 2. That is, $W_1(n_1^*) > W_2(n_2^*)$.
3. The total wait function $W_1(n)$ is strictly increasing at $n = N \times \frac{\lambda'_1}{\lambda'_1 + \lambda'_2}$.
4. At the allocation proportional to demand, the wait time in region 2 is strictly larger than that in region 1. That is, if we set $n_i = N \times \frac{\lambda'_i}{\lambda'_1 + \lambda'_2}$, then $W_2(n_2) > W_1(n_1)$.

Proof of Lemma A3. We start by statement 1. To see this, first note that from the assumption that (n_1^*, n_2^*) was the multi-region equilibrium under $(\lambda_1, \lambda_2, N, t)$, we know n_2^* has to be strictly larger than where the old W_2 function reached its trough. That is, $n_2^* > \sqrt{\lambda_2 t}$. Now, given $\lambda'_2 < \lambda_2$, we it is also the case that $n_2^* > \sqrt{\lambda'_2 t}$. Therefore, the new W_2 function is also strictly increasing at $n = n_2^*$.

We now turn to statement 2. Given that (n_1^*, n_2^*) was the multi-region equilibrium under the old parameters, the total wait times in the two regions were equal to each other. That is:

$$\frac{n_1^*}{\lambda_1} + \frac{t}{n_1^*} = \frac{n_2^*}{\lambda_2} + \frac{t}{n_2^*} \quad (16)$$

Given Proposition (4), we know that $n_1^* > n_2^*$, therefore: $\frac{t}{n_1^*} < \frac{t}{n_2^*}$. This latter inequality, combined with equality (16), implies $\frac{n_1^*}{\lambda_1} > \frac{n_2^*}{\lambda_2}$. The sign of this inequality is preserved if we multiply both sides of it by the positive number $\gamma - 1$. That is: $(\gamma - 1) \times \frac{n_1^*}{\lambda_1} > (\gamma - 1) \times \frac{n_2^*}{\lambda_2}$. The size of the inequality is also preserved when we add equal numbers to both sides. Those equal numbers are the two sides of equation (16). This will give us:

$$(\gamma - 1) \times \frac{n_1^*}{\lambda_1} + \frac{n_1^*}{\lambda_1} + \frac{t}{n_1^*} > (\gamma - 1) \times \frac{n_2^*}{\lambda_2} + \frac{n_2^*}{\lambda_2} + \frac{t}{n_2^*} \quad (17)$$

Therefore:

$$\gamma \times \frac{n_1^*}{\lambda_1} + \frac{t}{n_1^*} > \gamma \times \frac{n_2^*}{\lambda_2} + \frac{t}{n_2^*} \quad (18)$$

which gives us:

$$\frac{n_1^*}{\frac{\lambda_1}{\gamma}} + \frac{t}{n_1^*} > \frac{n_2^*}{\frac{\lambda_2}{\gamma}} + \frac{t}{n_2^*} \quad (19)$$

which, by definition, means:

$$\frac{n_1^*}{\lambda'_1} + \frac{t}{n_1^*} > \frac{n_2^*}{\lambda'_2} + \frac{t}{n_2^*} \quad (20)$$

This proves statement 2.

Next, we turn to statement 3. The argument is similar to that for statement 1. $N \times \frac{\lambda_1}{\lambda_1 + \lambda_2}$ was larger than the trough of W_1 under the old parameters. Given that the trough gets smaller under the new parameters, it will keep being smaller than $N \times \frac{\lambda_1}{\lambda_1 + \lambda_2}$.

Finally, statement 4 is obvious from the proof of Proposition (4). \square

Now notice that Lemma (A3) completes the proof of the proposition. Given that under the allocation (n_1^*, n_2^*) , we have $W_1 > W_2$, and under the allocation fully proportional to demand, we have $W_1 < W_2$, and given that both W_1 and W_2 are continuous functions, there should be an allocation $(n_1^{*'}, n_2^{*''})$ in between the two such that $W_1(n_1^{*'}) = W_2(n_2^{*''})$. This was achieved by

statements 2 and 4. Now by statement 1, W_2 is strictly increasing at $n = n_2^{*'} > n_2^*$ because $n_2^{*'} > n_2^* > \sqrt{\lambda'_2 t}$. Also, by statement 3, W_1 is increasing at $n = n_1^{*'} > n_1^*$ because $n_1^{*'} > N \frac{\lambda_1}{\lambda_1 + \lambda_2} > \sqrt{\lambda'_1 t}$. This implies that $(n_1^{*'}, n_2^{*'})$ is the multi-region equilibrium under parameters $(\lambda'_1, \lambda'_2, N, t)$. Now, given $n_1^{*'} < n_1^*$ and $n_2^{*'} > n_2^*$, it follows that:

$$\kappa^{*'} = \frac{\frac{n_1^{*'}}{\lambda'_1}}{\frac{n_2^{*'}}{\lambda'_2}} < \frac{\frac{n_1^*}{\lambda'_1}}{\frac{n_2^*}{\lambda'_2}} = \frac{\frac{n_1^*}{n_2^*}}{\frac{\lambda_1}{\lambda_2}} = \kappa^*$$

which finishes the proof of the proposition. ■

Proof of Proposition (7). Before stating the induction hypothesis, we add one statement to the five statements of Proposition (7). The inclusion of this statement and leveraging it in the induction process will be helpful for the proof. We call it statement 6.

Statement 6. Suppose an all region equilibrium $n^* = (n_1^*, \dots, n_I^*)$ exists under primitives (λ, N, t) where $\lambda = (\lambda_1, \dots, \lambda_I)$. Then, demand arrival rates are scaled down, that is, under new primitives $(\frac{\lambda}{\gamma}, N, t)$ with $\gamma > 1$, we have:

- An all-regions equilibrium $n^{*'} = (n_1^{*'}, \dots, n_I^{*'})$ exists.
- The new equilibrium $n^{*'}$ shows less geographical supply inequity than n^* in the sense that for any $i < j$, we have $\frac{\frac{n_i^{*'}}{\lambda_i}}{\frac{n_j^{*'}}{\lambda_j}} \leq \frac{\frac{n_i^*}{\lambda_i}}{\frac{n_j^*}{\lambda_j}}$. The inequality is strict if and only if $\lambda_i > \lambda_j$.

In words, this statement simply says the geographical supply inequity decreases if, all else fixed, all demand arrival rates proportionally decrease. The intuition is that this makes idle times relatively more important than pickup times.

We can now state the strong induction hypothesis.

Induction Hypothesis. Take some natural number $I_0 > 2$. If all statements of Proposition (7), including statement 6 added above, are correct for $I \in \{2, \dots, I_0 - 1\}$, then they are also all correct for $I = I_0$.

Now, in order to prove the proposition, we need to take two steps. First, we should prove the basis of the induction process. That is, we must show the proposition holds under $I = 2$. Second, we need to prove the induction hypothesis. As for the first step, note that propositions (2) through (6) do this job. The only statement that is not explicitly proven by those theorem is statement 6. However, the proof of statement 6 was the main building block of the proof of Proposition (6).¹⁹

We now turn to the second and main step of this proof, which is to show that the induction hypothesis is correct (Note that some of the statements are not really proven based on the induction.

¹⁹Also, propositions (2) through (6) assume that $\lambda_1 > \lambda_2$ and, hence, leave out the case where $\lambda_1 = \lambda_2$. But the proofs for the case where $I = 2$ and $\lambda_1 = \lambda_2$ are straightforward and we leave them to the reader.

Nevertheless, we present all of the proofs in this inductive framework since we believe having one induction as well as one non-induction section for the proof will just make it harder to read).

Proof of Statement 1. If the total wait time in region i is strictly higher than that in region j , then given the continuity of these wait-time functions, a small enough mass of drivers can leave region i for j and strictly benefit from that, violating the equilibrium assumption. To see why they are increasing, suppose on the contrary, that at the equilibrium allocation, for region i , the total wait time is strictly decreasing in the number of drivers in that region. Since drivers are equal across all regions in equilibrium, drivers from any other region j will have the incentive to relocate to region i , given that (i) currently region i has the same total wait as they do; and (ii) once they move to region i , the total wait time of that region will decrease. This is a violation of the equilibrium assumption. Therefore it has to be that at the equilibrium, the wait times are all increasing in the number of drivers at all regions. \square

Proof of Statement 2. Suppose, on the contrary, that there are two different all-regions equilibria n^* and \bar{n} . Given the two vectors are different, there has to be a region i such that $n_i^* \neq \bar{n}_i$. Without loss of generality, assume $n_i^* < \bar{n}_i$. Given that, from statement 1, we know the total wait time is increasing at n_i^* , and given the fact that the wait time function, once it becomes increasing, it remains strictly increasing, we can say $W_i(n_i^*) < W_i(\bar{n}_i)$.

Now, again from statement 1, we know two things. First, $\forall j : W_j(n_j^*) = W_j(n_i^*) \quad \& \quad W_j(\bar{n}_j) = W_j(\bar{n}_i)$, which implies: $\forall j : W_j(n_j^*) < W_j(\bar{n}_j)$. Second, we know that the total wait time function at each region j must be strictly increasing after it hits its trough (which happens weakly before n_j^*). This implies that in order for $\forall j : W_j(n_j^*) < W_j(\bar{n}_j)$ to hold, it must be that $\forall j : n_j^* < \bar{n}_j$. Therefore:

$$\sum_{j=1, \dots, I_0} \bar{n}_j > \sum_{j=1, \dots, I_0} n_j^*$$

But this cannot be given that both of the sums should be equal to N . \square

Proof of Statement 3. Note that the definition of equilibrium is that no driver should have the incentive to relocate from one region to another. This definition, by construction, implies that if $n^* = (n_1^*, \dots, n_{I_0}^*)$ is an equilibrium under (λ, N, t) , then once we fix $\tilde{N} = n_i^* + n_j^*$ for some i, j with $i < j$, then the allocation (n_i^*, n_j^*) is itself an equilibrium of the two-region game with primitives $(\lambda_i, \lambda_j, \tilde{N}, t)$. Thus, by Proposition (4) (or alternatively, by the base of the induction), we know that if $\lambda_i > \lambda_j$, then $\frac{n_i^*}{\lambda_i} > \frac{n_j^*}{\lambda_j}$. Also in case $\lambda_i = \lambda_j$, it is fairly straightforward to verify that $\frac{n_i^*}{\lambda_i} = \frac{n_j^*}{\lambda_j}$. To see this, note that in that case, $\frac{n_i^*}{\lambda_i} = \frac{n_j^*}{\lambda_j}$ if and only if $n_i^* = n_j^*$. It is easy to see that $n_i^* = n_j^*$ is an equilibrium given that it gives the two regions the same total wait time and that at it, the total wait times must be increasing according to previous statements. \square

Proof of Statement 4. Before we start the proof of this statement, we note that, similar to the case of Proposition (5), we can work with primitives $(\lambda, N, \frac{t}{\gamma})$ instead of $(\gamma\lambda, \gamma N, t)$. As a reminder, this is because there is a one-to-one and onto mapping between the equilibria under the

two primitives, which preserves all of the $\frac{n_i^*}{\lambda_i}$ values.

We start by proving the first statement. That is, if an all-regions equilibrium exists under (λ, N, t) , then one does under $(\lambda, N, \frac{t}{\gamma})$ as well. To see this, let us assume that under the “old” primitives (λ, N, t) , the all-regions equilibrium allocation n^* is such that $\forall i \in \{1, \dots, I_0\} : W_i(n_i^*) = w$. We know this common w must exist from statement 1, and we know it is unique from statement 2. We show existence of an equilibrium allocation under the new primitive by first describing two “partial equilibrium” allocations. We construct the first partial equilibrium allocation $\bar{n} = (\bar{n}_1, \dots, \bar{n}_{I_0})$ by fixing $\bar{n}_1 = n_1^*$ and assuming the rest of values $(\bar{n}_2, \dots, \bar{n}_{I_0})$ to be the equilibrium allocation of drivers among regions 2 to I_0 under primitives $((\lambda_2, \dots, \lambda_{I_0}), N - n_1^*, \frac{t}{\gamma})$. In words, this allocation fixes the number of drivers in region 1 (i.e., the region with the highest demand arrival rate λ_1) at its value under the old primitives but allows the drivers of all other regions to reshuffle themselves among those regions. The second partial equilibrium allocation \tilde{n} fixes $\tilde{n}_1 = N \times \frac{\lambda_1}{\sum_{i=1, \dots, I_0} \lambda_i}$, and assumes the rest of values $(\tilde{n}_2, \dots, \tilde{n}_{I_0})$ to be the equilibrium allocation of drivers among regions 2 to I_0 under primitives $((\lambda_2, \dots, \lambda_{I_0}), N \times (1 - \frac{\lambda_1}{\sum_{i=1, \dots, I_0} \lambda_i}), \frac{t}{\gamma})$. In words, this allocation fixes the total number of drivers in region 1 at the value it would take if drivers were to be allocated fully proportional to demand arrival rates. It then allows the rest of the drivers to reshuffle themselves among other areas under the new primitives. We will use these two partial equilibrium allocations to prove existence of an all-region equilibrium. But first we need to prove the existence of these partial equilibrium allocations themselves. Lemma A4 below does this job.

Lemma A4. *Partial equilibrium allocations \tilde{n} and \bar{n} described above exist, are unique, and allocate a strictly positive number of drivers to each region.*

Proof of Lemma (A4). We first start from \bar{n} . Note that the assumption of n^* being the equilibrium allocation under (λ, N, t) , by construction implies that $(n_2^*, \dots, n_{I_0}^*)$ is the unique all-region equilibrium allocation under $((\lambda_2, \dots, \lambda_{I_0}), N - n_1^*, t)$. Now, given that by our induction assumption all results (including statement 4) hold for $I_0 - 1$ regions, if the primitives remain the same except that t is divided by some $\gamma > 1$, a unique all-region equilibrium will still exist. This is what we were denoting \bar{n}_2 through \bar{n}_{I_0} .

Next, we turn to \tilde{n} and construct it from \bar{n} . We just showed that $(\bar{n}_2, \dots, \bar{n}_{I_0})$ is the unique all-regions equilibrium under $((\lambda_2, \dots, \lambda_{I_0}), N - n_1^*, t)$. Also note that by statement 3, we know $n_1^* > \frac{\lambda_1}{\sum_{i=1, \dots, I_0} \lambda_i}$, which implies $N - n_1^* < N \times (1 - \frac{\lambda_1}{\sum_{i=1, \dots, I_0} \lambda_i})$. Therefore, primitives $((\lambda_2, \dots, \lambda_{I_0}), N \times (1 - \frac{\lambda_1}{\sum_{i=1, \dots, I_0} \lambda_i}), \frac{t}{\gamma})$ can be constructed from primitives $((\lambda_2, \dots, \lambda_{I_0}), N - n_1^*, t)$ by increasing the total number of drivers. Given that \bar{n} was the unique all-regions equilibrium allocation under $((\lambda_2, \dots, \lambda_{I_0}), N - n_1^*, t)$, and given the induction assumption on statement 5 for $I = I_0 - 1$ regions, we can say that primitives $((\lambda_2, \dots, \lambda_{I_0}), N \times (1 - \frac{\lambda_1}{\sum_{i=1, \dots, I_0} \lambda_i}), \frac{t}{\gamma})$ also have a unique all-regions equilibrium allocation. This is exactly what was denoted $\tilde{n}_2, \dots, \tilde{n}_{I_0}$. This completes the proof of the lemma. \square

We now use these two partial equilibrium allocations to show that a unique all-regions equilibrium allocation exists under primitives $(\lambda, N, \frac{t}{\gamma})$. Our next step is to prove the following useful lemma.

Lemma A5. *At the partial equilibrium allocation \bar{n} , the total wait time in region 1 is larger than that in any other region. Conversely, at the partial equilibrium allocation \tilde{n} , the total wait time in region 1 is smaller than that in any other region.*

Proof of Lemma (A5). To see why the result holds for \bar{n} , note that under the old equilibrium n^* and old primitives (λ, N, t) , all of the wait times were equal. This means for any $i > 1$ we had

$$\frac{n_1^*}{\lambda_1} + \frac{t}{n_1^*} = \frac{n_i^*}{\lambda_i} + \frac{t}{n_i^*}$$

But given that for all $i > 1$ we have $n_1^* \geq n_i^*$ we get the following inequality under the new primitives $(\lambda, N, \frac{t}{\gamma})$:

$$\frac{n_1^*}{\lambda_1} + \frac{\frac{t}{\gamma}}{n_1^*} \geq \frac{n_i^*}{\lambda_i} + \frac{\frac{t}{\gamma}}{n_i^*}$$

Next, note that the main and only difference between allocations n^* and \bar{n} is that under \bar{n} , drivers reshuffle among regions 2 to I_0 in order to reduce their total wait times. Therefore, there has to be at least one region j such that:

$$\frac{\bar{n}_j}{\lambda_j} + \frac{\frac{t}{\gamma}}{\bar{n}_j} \leq \frac{n_j^*}{\lambda_j} + \frac{\frac{t}{\gamma}}{n_j^*}$$

Combining the above two, we get:

$$\frac{\bar{n}_j}{\lambda_j} + \frac{\frac{t}{\gamma}}{\bar{n}_j} \leq \frac{n_1^*}{\lambda_j} + \frac{\frac{t}{\gamma}}{n_1^*}$$

But the total wait time under $(\lambda, N, \frac{t}{\gamma})$ is equal across regions 2 through I_0 under allocation \bar{n} . Therefore, the above inequality holds not only for a specific j , but under any $j > 1$. This proves the lemma for \bar{n} given that $\bar{n}_1 = n_1^*$.

Next, we prove the lemma for \tilde{n} . We first show that the wait time in region 1 is smaller than that in region 2 if both get drivers proportional to their demand arrival rates. We then show that the wait time in region 2 under \tilde{n}_2 is larger than the wait time in region 2 if region 2 were to get drivers proportional to its demand arrival rate. These two statements, combined, will prove our intended result. To see the first claim, note that the wait time in region 1, if it gets $N \times \frac{\lambda_1}{\sum_{i \in \{1, \dots, I_0\}} \lambda_i}$ drivers, will be:

$$w_1 = \frac{N \times \frac{\lambda_1}{\sum_{i \in \{1, \dots, I_0\}} \lambda_i}}{\lambda_1} + \frac{\frac{t}{\gamma}}{N \times \frac{\lambda_1}{\sum_{i \in \{1, \dots, I_0\}} \lambda_i}}$$

which gives:

$$w_1 = \frac{N}{\sum_{i \in \{1, \dots, I_0\}} \lambda_i} + \frac{\frac{t}{\gamma} \times \sum_{i \in \{1, \dots, I_0\}} \lambda_i}{N \lambda_1} \quad (21)$$

Similarly, if region 2 were to get $N \times \frac{\lambda_2}{\sum_{i \in \{1, \dots, I_0\}} \lambda_i}$ drivers, its total wait time will be:

$$w_2 = \frac{N}{\sum_{i \in \{1, \dots, I_0\}} \lambda_i} + \frac{\frac{t}{\gamma} \times \sum_{i \in \{1, \dots, I_0\}} \lambda_i}{N \lambda_2} \quad (22)$$

It is easy to see that the first terms of w_1 and w_2 are the same, and the second term is larger in w_2 given that $\lambda_1 \geq \lambda_2$. Now note that under allocation \tilde{n} , the wait time in region 1 is indeed w_1 . So, it remains to show that $W_2(\tilde{n}_2) \geq w_2$. To show this, we make two observations (and prove them both shortly). First, $\tilde{n}_2 \geq N \times \frac{\lambda_2}{\sum_{i \in \{1, \dots, I_0\}} \lambda_i}$. This simply says under \tilde{n} , region 2 is getting more drivers than it would if drivers were to be allocated to regions proportionally to their demand rates. Second, the total wait time function in region 2 is increasing between $N \times \frac{\lambda_2}{\sum_{i \in \{1, \dots, I_0\}} \lambda_i}$ and \tilde{n}_2 . Together, these two observations imply $W_2(\tilde{n}_2) \geq w_2$, as desired. Therefore, we have shown that $W_2(\tilde{n}_2) \geq w_1$. But given that $(\tilde{n}_2, \dots, \tilde{n}_{I_0})$ was an all-regions equilibrium under $((\lambda_2, \dots, \lambda_{I_0}), N \times (1 - \frac{\lambda_1}{\sum_{i=1, \dots, I_0} \lambda_i}, \frac{t}{\gamma}))$, we know that for any $i, j > 1$: $W_i(\tilde{n}_i) = W_j(\tilde{n}_j)$. This, combined with $W_2(\tilde{n}_2) \geq w_1$, completes the proof of the lemma, of course with the exception of the two observations made in this paragraph. We now turn to proving those observations and finish the proof of the lemma.

The first observation was that $\tilde{n}_2 \geq N \times \frac{\lambda_2}{\sum_{i \in \{1, \dots, I_0\}} \lambda_i}$. To see why this is true, note that \tilde{n} is the all-regions equilibrium under $((\lambda_2, \dots, \lambda_{I_0}), N \times (1 - \frac{\lambda_1}{\sum_{i=1, \dots, I_0} \lambda_i}, \frac{t}{\gamma}))$. Therefore, by our induction assumption on statement 3, region 2 will get disproportionately more drivers relative to all other regions, because it has the highest λ_i amongst regions $2, \dots, I_0$. That is $\forall i > 2$: $\frac{\tilde{n}_2}{\lambda_2} \geq \frac{\tilde{n}_i}{\lambda_i}$. It is then easy to show that:

$$\frac{\tilde{n}_2}{\lambda_2} \geq \frac{\sum_{i=2, \dots, I_0} \tilde{n}_i}{\sum_{i=2, \dots, I_0} \lambda_i} \quad (23)$$

But we know, from the primitives, that $\sum_{i=2, \dots, I_0} \tilde{n}_i = N \times (1 - \frac{\lambda_1}{\sum_{i=1, \dots, I_0} \lambda_i}) = N \times \frac{\sum_{i=2, \dots, I_0} \lambda_i}{\sum_{i=1, \dots, I_0} \lambda_i}$. Now, plugging this into (23) and rearranging, we get $\tilde{n}_2 \geq N \times \frac{\lambda_2}{\sum_{i \in \{1, \dots, I_0\}} \lambda_i}$, which is exactly our first observation.

We now turn to the proof of the second observation. That is, we want to show that the total wait time function in region 2 is increasing between $N \times \frac{\lambda_2}{\sum_{i \in \{1, \dots, I_0\}} \lambda_i}$ and \tilde{n}_2 . To see this, note that the wait time curve in region 2 takes the form that was depicted in figure (6). In particular, it is a curve with only one trough; and once past the trough, the curve will remain strictly increasing indefinitely. Thus, to prove that the wait-time is increasing over the interval $[N \times \frac{\lambda_2}{\sum_{i \in \{1, \dots, I_0\}} \lambda_i}, \tilde{n}_2]$, it is sufficient to show that the smallest point in this interval is past the trough. One can show the trough happens at $n_2 = \sqrt{\frac{t}{\gamma} \lambda_2}$. Therefore, what we need to show is:

$$N \times \frac{\lambda_2}{\sum_{i \in \{1, \dots, I_0\}} \lambda_i} \geq \sqrt{\frac{t}{\gamma} \lambda_2} \quad (24)$$

In order to prove this, we first assume, to the contrary, that $N \times \frac{\lambda_2}{\sum_{i \in \{1, \dots, I_0\}} \lambda_i} < \sqrt{\frac{t}{\gamma} \lambda_2}$; then we arrive at a contradiction with the result that \tilde{n} is the all-regions equilibrium under $((\lambda_2, \dots, \lambda_{I_0}), N \times (1 - \frac{\lambda_1}{\sum_{i=1, \dots, I_0} \lambda_i}, \frac{t}{\gamma}))$. Note that we are assuming, without loss of generality, $\lambda_2 \geq \lambda_i$ for any $i > 2$.

Therefore, given that all λ_i are positive, for any $i > 2$, we have $\frac{\lambda_i}{\lambda_2} \leq \sqrt{\frac{\lambda_i}{\lambda_2}}$. Thus, if we multiply the left hand side of the inequality $N \times \frac{\lambda_2}{\sum_{i \in \{1, \dots, I_0\}} \lambda_i} < \sqrt{\frac{t}{\gamma} \lambda_2}$ by $\frac{\lambda_i}{\lambda_2}$ and the right hand side by $\sqrt{\frac{\lambda_i}{\lambda_2}}$, then the direction of the inequality should not change. Therefore, not only for region 2, but also for any region $i \geq 2$, we will have:

$$N \times \frac{\lambda_i}{\sum_{j \in \{1, \dots, I_0\}} \lambda_j} < \sqrt{\frac{t}{\gamma} \lambda_i}$$

Now, if we sum over all $i = 2, \dots, I_0$ on both sides of the inequality above, we get:

$$N \times \sum_{i=2, \dots, I_0} \frac{\lambda_i}{\sum_{j \in \{1, \dots, I_0\}} \lambda_j} < \sum_{i=2, \dots, I_0} \sqrt{\frac{t}{\gamma} \lambda_i}$$

Rearranging, we get:

$$N \times \left(1 - \frac{\lambda_1}{\sum_{j \in \{1, \dots, I_0\}} \lambda_j}\right) < \sum_{i=2, \dots, I_0} \sqrt{\frac{t}{\gamma} \lambda_i}$$

But $N \times \left(1 - \frac{\lambda_1}{\sum_{j \in \{1, \dots, I_0\}} \lambda_j}\right)$ is the total number of drivers in regions 2 through I_0 . That is: $N \times \left(1 - \frac{\lambda_1}{\sum_{j \in \{1, \dots, I_0\}} \lambda_j}\right) = \sum_{j=2, \dots, I_0} \tilde{n}_j$. Therefore, we get:

$$\sum_{j=2, \dots, I_0} \tilde{n}_j < \sum_{i=2, \dots, I_0} \sqrt{\frac{t}{\gamma} \lambda_i}$$

which implies there should be at least one $j \geq 2$ such that $\tilde{n}_j < \sqrt{\frac{t}{\gamma} \lambda_j}$. But this means that for that region j , the wait time function is decreasing at $n = \tilde{n}_j$ contradicting the result that \tilde{n}_j is part of an all-regions equilibrium. This completes the proof of the second observation, and hence that of lemma (A5). \square

Next, we use lemma (A5) to construct an all-regions equilibrium under primitives $(\lambda, N, \frac{t}{\gamma})$. This will be a constructive proof to the existence portion of statement 4. To this end, we start from the first partial equilibrium \bar{n} , gradually shifting drivers from region 1 to other regions until we are left with \tilde{n} drivers in region 1. That is, for any $\hat{n}_1 \in [\tilde{n}_1, \bar{n}_1]$ we consider the partial equilibrium $\hat{n} = (\hat{n}_1, \dots, \hat{n}_{I_0})$ such that the $(\hat{n}_2, \dots, \hat{n}_{I_0})$ is the all-regions equilibrium allocation under primitives $(\lambda_2, \dots, \lambda_{I_0}), N - \hat{n}_1, \frac{t}{\gamma}$. The argument for why such partial equilibrium exists for any $\hat{n}_1 < \bar{n}_1$ is similar the argument given in proof of lemma (A4) for \tilde{n}_1 .

Now, note that by lemma (A5), the total wait time in region 1 is larger than that in other regions when $\hat{n}_1 = \bar{n}_1$ and it is smaller in region 1 than it is in other regions when $\hat{n}_1 = \tilde{n}_1$. Therefore, there should be some $\hat{n}_1 \in [\tilde{n}_1, \bar{n}_1]$ for which the total wait time in region 1 is equal to the total wait time in all of the other regions, which themselves are equal to each other by \hat{n} being a partial equilibrium the way defined above.²⁰ We claim such allocation \hat{n} is the all-regions equilibrium of

²⁰Note that in order to make this argument we also need to know that as we move \hat{n}_1 within $[\tilde{n}_1, \bar{n}_1]$, the total wait time in region 1 as well as the common total wait time in the other regions both move continuously. This is true by construction for region 1, since the total wait time function is continuous. For other regions, this needs to be shown that as we add drivers to the collection of these regions, the equilibrium total wait time moves continuously. We skip the proof of this claim here, but can provide it upon request.

the whole market (that is, under primitives $(\lambda, N, \frac{t}{\gamma})$). The proof for this claim is as follows:

We know that under allocation \hat{n} all regions have the same total wait time. We also know, by $(\hat{n}_2, \dots, \hat{n}_{I_0})$ being the all-regions equilibrium under primitives $(\lambda_2, \dots, \lambda_{I_0}, N - \hat{n}_1, \frac{t}{\gamma})$, that the total wait time in each region $i > 1$ is increasing at $n = \hat{n}_i$. Thus, the only thing that remains to be shown is that for $i = 1$ too the total wait time curve is increasing at $n = \hat{n}_1$. To this end, as argued before in a similar case, we need to show that $\hat{n}_1 \geq \sqrt{\frac{t}{\gamma} \lambda_1}$. Note that given $\hat{n}_1 \geq \tilde{n}_1$, it would suffice to show $\tilde{n}_1 \geq \sqrt{\frac{t}{\gamma} \lambda_1}$. We show this latter inequality by borrowing from what we already did in the proof of the last observation we made as part of proof of lemma (A5). There, we proved inequality (24) holds. Now, given that we have been assuming (without loss of generality) that $\lambda_1 \geq \lambda_2$, and given that all λ_i are positive numbers, we get: $\frac{\lambda_1}{\lambda_2} \geq \sqrt{\frac{\lambda_1}{\lambda_2}}$. Therefore, if we multiply the left-hand side of equation (24) by $\frac{\lambda_1}{\lambda_2}$ and the right hand side by $\sqrt{\frac{\lambda_1}{\lambda_2}}$, the sign of the inequality should not change. This operation gets us:

$$N \times \frac{\lambda_1}{\sum_{i \in \{1, \dots, I_0\}} \lambda_i} \geq \sqrt{\frac{t}{\gamma} \lambda_1}$$

which is exactly what we were after. This shows that \hat{n} is the all-regions equilibrium, completing the proof of the first part of statement 4 in the proposition. \square

Now that we have shown the all-region equilibrium \hat{n} under primitives $(\lambda, N, \frac{t}{\gamma})$ exists, we show that it indeed shows less geographical supply inequity than the old equilibrium n^* . As the first step towards this goal, note that for any $j > i > 1$, we can show the result holds based on our induction assumption. More precisely, we know that $(n_2^*, \dots, n_{I_0}^*)$ is the all-regions equilibrium under primitives $((\lambda_2, \dots, \lambda_{I_0}), N - n_1^*, t)$. We also know that $(\hat{n}_2, \dots, \hat{n}_{I_0})$ is the all regions equilibrium under primitives $((\lambda_2, \dots, \lambda_{I_0}), N - \hat{n}_1, \frac{t}{\gamma})$. The move from primitives $((\lambda_2, \dots, \lambda_{I_0}), N - n_1^*, t)$ to primitives $((\lambda_2, \dots, \lambda_{I_0}), N - \hat{n}_1, \frac{t}{\gamma})$ involves two steps. The first step is to divide t by some $\gamma > 1$. The second step is to add $n_1^* - \hat{n}_1$ drivers. Based on our induction assumption, both statements 4 and 5 of Proposition (7) hold for $I_0 - 1$ regions. Therefore, for any $j > i > 1$ we have:

$$\frac{\frac{\hat{n}_i}{\lambda_i}}{\frac{\hat{n}_j}{\lambda_j}} \leq \frac{\frac{n_i^*}{\lambda_i}}{\frac{n_j^*}{\lambda_j}}$$

with the inequality strict if $\lambda_i > \lambda_j$. Now the only thing that remains to show is that we can say the same not only for $j > i > 1$, but also for $j > i = 1$. In order to show this, we consider three cases.

Case 1: for every $j > 1$, we have $\hat{n}_j > n_j^*$. In this case, the result is becomes trivial given that we know $\hat{n}_1 \leq n_1^*$.

Case 2: for at least two distinct $j, j' > 1$, we have $\hat{n}_j \leq n_j^*$ and $\hat{n}_{j'} \leq n_{j'}^*$. We start with j and note that the allocation of drivers in all regions other than j –i.e., allocation $(\hat{n}_1, \dots, \hat{n}_{j-1}, \hat{n}_{j+1}, \dots, \hat{n}_{I_0})$ is the all-region equilibrium under primitives $((\lambda_1, \dots, \lambda_{j-1}, \lambda_{j+1}, \dots, \lambda_{I_0}), N -$

$\hat{n}_j, \frac{t}{\gamma}$). Also note that allocation $(n_1^*, \dots, n_{j-1}^*, n_{j+1}^*, \dots, n_{I_0}^*)$ is the all-region equilibrium under primitives $((\lambda_1, \dots, \lambda_{j-1}, \lambda_{j+1}, \dots, \lambda_{I_0}), N - n_j^*, \frac{t}{\gamma})$. Note that the former primitives can be obtained from the latter by two moves. First, going from t to $\frac{t}{\gamma}$ for some $\gamma > 1$; and second, changing the total number of drivers from $N - n_j^*$ to the (by assumption) larger number of $N - \hat{n}_j$. Based on our induction assumptions, we know that both of these moves reduce the geographical supply inequity. Therefore, now we can claim the following for any $i \neq j$:

$$\frac{\frac{\hat{n}_1}{\lambda_1}}{\frac{\hat{n}_i}{\lambda_i}} \leq \frac{\frac{n_1^*}{\lambda_1}}{\frac{n_i^*}{\lambda_i}}$$

with the inequality strict if $\lambda_1 \neq \lambda_i$. This covers all of the comparisons that we needed with the exception of the comparison between region 1 and region j itself. But we can prove the inequality for that case as well, by going through the exact same process as above, except excluding region j' this time instead of region j . This finishes the proof of statement 4 of the proposition under case 2.

Case 3: for exactly one region $j > 1$, we have $\hat{n}_j \leq n_j^*$. In this case, we can go through the same process as that described in case 2, to show for any $i \neq j$:

$$\frac{\frac{\hat{n}_1}{\lambda_1}}{\frac{\hat{n}_i}{\lambda_i}} \leq \frac{\frac{n_1^*}{\lambda_1}}{\frac{n_i^*}{\lambda_i}}$$

with the inequality strict if $\lambda_1 \neq \lambda_i$. This time, however, we are not able to use a similar argument for the to show the result holds between regions 1 and j . The following lemmas, however, demonstrate a different way to prove the result for this specific comparison.

Lemma A6. *Under the conditions of case 3, we have $\hat{n}_1 \geq \frac{n_1^*}{\sqrt{\gamma}}$ and $\hat{n}_j \geq \frac{n_j^*}{\sqrt{\gamma}}$.*

Proof of Lemma (A6). We only show $\hat{n}_1 \geq \frac{n_1^*}{\sqrt{\gamma}}$. The argument for $\hat{n}_j \geq \frac{n_j^*}{\sqrt{\gamma}}$ is the same.

We start by observing that the total wait time w_1 in region 1 under $n_1 = \frac{n_1^*}{\sqrt{\gamma}}$ is given by:

$$\begin{aligned} w_1 &= \frac{n_1}{\lambda_1} + \frac{\frac{t}{\gamma}}{n_1} \\ &= \frac{\frac{n_1^*}{\sqrt{\gamma}}}{\lambda_1} + \frac{\frac{t}{\gamma}}{\frac{n_1^*}{\sqrt{\gamma}}} \\ &= \left(\frac{n_1^*}{\lambda_1} + \frac{t}{n_1^*} \right) \frac{1}{\sqrt{\gamma}} \\ &= \frac{w^*}{\sqrt{\gamma}} \end{aligned} \tag{25}$$

where w^* is the common total wait time among all regions under primitives (λ, N, t) and the all-regions equilibrium n^* given those primitives.

Next, we show that under the *new primitives* $(\lambda, N, \frac{t}{\gamma})$, but at the *old equilibrium allocation* n^* , the total wait-time in any region i is weakly larger than $\frac{w^*}{\sqrt{\gamma}}$. To see this, we write out one such total wait time:

$$\frac{n_i^*}{\lambda_i} + \frac{\frac{t}{\gamma}}{n_i^*}$$

Note that because n^* is the all-region equilibrium under the old primitives, it must be that for all i : $n_i^* \geq \sqrt{t\lambda_i}$. This gives $\frac{n_i^*}{\lambda_i} \geq \frac{t}{n_i^*}$, or, alternatively: $\frac{t}{n_i^*} \leq \frac{1}{2}(\frac{n_i^*}{\lambda_i} + \frac{t}{n_i^*}) = \frac{w^*}{2}$.

Therefore, we can write:

$$\begin{aligned} \frac{n_i^*}{\lambda_i} + \frac{\frac{t}{\gamma}}{n_i^*} &= (\frac{n_i^*}{\lambda_i} + \frac{t}{n_i^*}) - \frac{t}{n_i^*}(1 - \frac{1}{\gamma}) \\ &= w^* - \frac{t}{n_i^*}(1 - \frac{1}{\gamma}) \\ &\geq w^*(1 - \frac{1}{2}(1 - \frac{1}{\gamma})) \\ &= w^* \times (\frac{1 + \frac{1}{\gamma}}{2}) \\ &> w^* \times (\sqrt{1 \times \frac{1}{\gamma}}) \\ &= \frac{w^*}{\sqrt{\gamma}} \end{aligned} \tag{26}$$

Equations (25) and (26), together, tell us that for any i , we have $\frac{n_i^*}{\lambda_i} + \frac{\frac{t}{\gamma}}{n_i^*} > w_1$. Now notice that the total wait time under the new primitives at the old equilibrium allocation in any region is increasing. This is simply because $\forall i$: $n_i^* \geq \sqrt{t\lambda_i} > \sqrt{\frac{t}{\gamma}\lambda_i}$. This, combined with the fact that there is at least one region i with $\hat{n}_i > n_i^*$,²¹ tells us:

$$\hat{w} \equiv \frac{\hat{n}_i}{\lambda_i} + \frac{\frac{t}{\gamma}}{\hat{n}_i} > \frac{n_i^*}{\lambda_i} + \frac{\frac{t}{\gamma}}{n_i^*} > w_1$$

where \hat{w} is defined as the common total wait time among all regions under the new primitives and new equilibrium allocation.

What $\hat{w} > w_1$ tells us is that if we reduce the number of drivers in region 1 to $n_1 = \frac{n_1^*}{\sqrt{\gamma}}$, the total wait time in region 1 falls below the equilibrium total wait time. But this means it has to be that $\hat{n}_1 > n_1 = \frac{n_1^*}{\sqrt{\gamma}}$. To see why, consider two scenarios. First, if $n_1 < \sqrt{\frac{t}{\gamma}\lambda_1}$, then by $\hat{n}_1 \geq \sqrt{\frac{t}{\gamma}\lambda_1}$, we get $\hat{n}_1 > n_1$. Next, if $n_1 \geq \sqrt{\frac{t}{\gamma}\lambda_1}$, then the wait time curve is strictly increasing when moving up from n_1 , which means at some point past n_1 , it hits the higher wait time $\hat{w} > w_1$. That point would be \hat{n}_1 . Thus, the lemma has been proven for region 1. The proof for region j is exactly the same. \square

²¹This is true because there are at least three regions; and besides regions 1 and j , case 3 assumes $\hat{n}_i > n_i^*$ for all i .

We now present the another useful lemma which helps us better understand what happens to the two regions 1 and j .

Lemma A7. *Consider a market with two regions 1 and 2 only. Allocation n^* is an equilibrium in this market under primitives $(\lambda_1, \lambda_2, N, t)$ if and only if allocation $\frac{n^*}{\sqrt{\gamma}}$ is an equilibrium under primitives $(\lambda_1, \lambda_2, \frac{N}{\sqrt{\gamma}}, \frac{t}{\gamma})$.*

Proof of Lemma (A7). Follows directly from definitions. \square

Now, lemmas (A6) and (A7) show us a clear way to complete the last piece of the inductive proof of statement 4 in the proposition. Based on lemma (A6), we know $\hat{n}_1 + \hat{n}_j > \frac{n_1^* + n_j^*}{\sqrt{\gamma}}$. Now define $N^* = n_1^* + n_j^*$ and $\hat{N} = \hat{n}_1 + \hat{n}_j$. We know that (n_1^*, n_j^*) was the all-regions equilibrium under primitives $(\lambda_1, \lambda_j, N^*, t)$. Thus, by lemma (A7) we can claim that $(\frac{n_1^*}{\sqrt{\gamma}}, \frac{n_j^*}{\sqrt{\gamma}})$ is the all-regions equilibrium under primitives $(\lambda_1, \lambda_j, \frac{N^*}{\sqrt{\gamma}}, \frac{t}{\gamma})$.

On the other hand, we know that (\hat{n}_1, \hat{n}_j) is the all-regions equilibrium under primitives $(\lambda_1, \lambda_j, \hat{N}, \frac{t}{\gamma})$. Given that we showed $\hat{N} > \frac{N^*}{\sqrt{\gamma}}$, and given that by our strong induction assumption statement 5 is correct for all two-region cases, we can write:

$$\frac{\frac{\hat{n}_1}{\lambda_1}}{\frac{\hat{n}_j}{\lambda_j}} \geq \frac{\frac{n_1^*}{\sqrt{\gamma}}}{\frac{n_j^*}{\sqrt{\gamma}}} = \frac{\frac{n_1^*}{\lambda_1}}{\frac{n_j^*}{\lambda_j}}$$

with the inequality strict whenever $\lambda_1 > \lambda_j$. This completes the proof of case 3, and hence finishes the inductive proof of statement 4 of Proposition (7). \square

Proof of Statement 6. The steps of this proof closely (almost exactly) follow the steps of the proof of statement 4. We skip it but can provide the detailed proof upon request.

Proof of Statement 5. Similar to the corresponding two-region case (i.e., proof of Proposition (6)). This statement can be proven in a straightforward manner once we have proven statements 4 and 6. To be more precise, if we know that geographical supply inequity decreases in the sense defined in the statement of the proposition both (i) when we proportionally scale-up N and the vector λ and (ii) when we scale down the vector λ , it follows that the geographical supply inequity also decreases when we only scale up N , which is a certain combination of (i) and (ii). \square

The above proofs show that (i) the proposition holds for $I = 2$ and that (ii) the proposition holds for any $I_0 > 2$ if it holds for all $I \in \{2, \dots, I_0 - 1\}$. This means our proof is complete. \blacksquare