

Enabling the Unavailable Trip: Consumer Welfare and Willingness to Pay for Multimodal Technology

Michela Bonani and Julian Hidalgo*

October 19, 2026

Abstract

We recover the valuation for multimodal technologies using detailed data from a long-distance transportation platform in France. About 34% of the trips booked on the platform start or end in a different location from the one consumers initially searched for. We estimate a structural demand model and quantify the disutility arising from these spatial mismatches. We find that consumers are willing to pay 6.95 EUR, about a third of a fare, to eliminate the average distance gap, and older and privacy-sensitive consumers would pay more than younger ones. We use our estimates to study the gains from eliminating spatial mismatches through the multimodal technology. The technology would raise total consumer surplus by 10.4 EUR million: 0.57 EUR per searcher or 5.6 EUR per booking. We then show that the platform could charge a uniform fee of up to 5.12 EUR per trip, about a quarter of the average fare, for the technology without reducing the average surplus.

Keywords: Transportation Markets, Demand Estimation, Digital Platform.

JEL Codes: D12, L91, R41, R42.

*M. Bonani is at Universitat Pompeu Fabra and J. Hidalgo is at Universitat Pompeu Fabra. M. Bonani: Department of Economics and Business, UPF. E-mail: michela.bonani@upf.edu J. Hidalgo: Department of Economics and Business, UPF. E-mail: julianfelipe.hidalgo@upf.edu.

1 Introduction

The European Commission has introduced several legislative and policy initiatives to promote a more integrated and multimodal transport sector, with the broader aim of reducing emissions and encouraging more sustainable travel ([European Parliament and Council of the European Union, 2010](#); [European Commission, 2017, 2023](#); [European Parliament, 2026](#)). Multimodal technologies can contribute to this objective by making it easier for passengers to combine different modes of transport, potentially reducing their reliance on private cars, including for the portions of a journey needed to reach or leave a public or shared transport service. However, combining transport modes is not always straightforward. Finding a suitable itinerary may require searching across different services, while the available options may not connect passengers' desired origins and destinations or offer convenient departure times. These frictions can limit the set of feasible journeys and make it harder for passengers to identify suitable alternatives. By integrating different transport modes within a single search and booking process, digital platforms can help address these problems. Multimodal integration can make journeys feasible by combining services that do not cover the entire route individually and expand the alternatives available to passengers on routes that are already served.

Despite growing policy interest in multimodal transport, empirical evidence on the need for these technologies and their potential benefits for consumers remains limited. How large are the frictions that prevent passengers from finding suitable itineraries? How much do these frictions affect their travel choices and welfare? And to what extent can multimodal integration improve consumer outcomes? This paper addresses these questions by documenting the frictions associated with spatial mismatches between searched and booked trips, estimating the disutility they impose on passengers, and quantifying the potential welfare gains from reducing them.

We study these questions using proprietary data from BlaBlaCar, a leading long-distance carpooling platform that also operates bus services on major corridors. Passengers searching for a trip on the platform may face two types of market frictions. First, the searched origin-destination pair may not be served by either carpooling or bus services, leaving passengers without a direct option for their desired journey. Second, even when a route is served, passengers may not find a suitable trip at the time of their search. This reflects the fact that drivers and bus operators post trips at different times, while passengers search according to their own travel needs. In both cases, passengers may leave the platform without booking or accept an alternative that does not exactly match their desired itinerary (a near-match booking). When they book such an alternative, they must cover the distance between their searched and booked origins and destinations themselves. This spatial gap provides a measure of the extent of the mismatch and

the associated disutility for passengers.

Multimodal technology could address these frictions by allowing passengers to combine transport modes within a single itinerary. First, combining carpooling and bus services can make origin-destination pairs feasible even when neither mode provides a suitable trip on its own. Second, it can expand the set of alternatives available to passengers searching routes that are already served, increasing the chances of finding an itinerary that better matches their travel needs. By connecting services across modes, multimodal integration could reduce the spatial gaps passengers need to cover themselves and make more trips feasible. It could therefore benefit both passengers who already book a trip but must accept a spatial mismatch and those who would otherwise leave the platform without booking.

We first document these market frictions and establish their magnitude. Using individual-level search and booking data from BlaBlaCar in France between 2021–2022, combined with information on the available supply of carpooling and bus services, we examine how often passengers book trips that differ from their requested itineraries, how far these trips depart from the requested origins and destinations, and how these patterns vary across transport modes and routes. We also distinguish between searches for which supply is unavailable and those for which trips are offered but passengers do not find a suitable option. This analysis establishes the extent of the problem and identifies the settings in which multimodal integration may be most valuable.

The descriptive evidence shows that spatial mismatches are common among passengers who book a trip: approximately 34% of bookings differ from the searched origin-destination pair. The gaps can also be substantial, averaging around 50 km. At the same time, only a small share of searches is classified as unserved. This suggests that the frictions documented in the data extend beyond cases in which no supply is available. Together, these patterns highlight the importance of distinguishing between the availability of transport services and the ability of passengers to find an itinerary that meets their needs.

To estimate the disutility associated with spatial mismatches, we develop an empirical model of transportation-mode choice among carpooling, buses, and an outside option. We model consumer choices within a static discrete-choice framework, following [Goolsbee and Petrin \(2004\)](#), and estimate the demand parameters using a multinomial probit model. We exploit variation in the distance between searched and booked itineraries, which arises because the platform’s search algorithm displays both exact-match and near-match trips. A near-match trip in one market may be an exact-match trip in another: it was posted to serve a different origin-destination pair but may be displayed to passengers searching elsewhere because it offers a potential substitute.

Variation in the gaps associated with displayed trips therefore partly reflects the availability of services intended for other routes, rather than only the supply of trips designed to serve the searched route.

This source of variation helps identify the disutility of spatial gaps under the assumption that, conditional on route and calendar fixed effects and other controls, demand shocks on the searched route are uncorrelated with demand shocks on the routes for which near-match trips were originally posted. The identifying argument also relies on the fact that drivers cannot target the markets in which their trips are listed as near matches: the platform determines which trips are displayed based on their proximity to the searched endpoints and other trip characteristics. These features make it less likely that the gap associated with a displayed trip reflects a driver’s response to demand in the market where the trip is ultimately shown. A remaining concern is that the average gap is computed from realized bookings and may therefore reflect the availability of exact-match trips or other local market conditions. We address this concern by controlling for the number of offers on the searched route and the number of results displayed to the passenger.

Since prices may be correlated with unobserved components of consumer utility, we use instrumental variables to identify price sensitivity. Specifically, following [De Loecker and Scott \(2026\)](#), we use measures based on fuel prices, travel distances, travel times, and mode-specific operating costs as instruments for prices.

The demand estimates show that spatial mismatches between searched and booked trips generate significant disutility for consumers. This effect is mainly driven by large gaps at the origin and destination, while consumers are largely indifferent to minor mismatches. Consumers are also willing to pay to reduce these gaps, with an average willingness to pay of EUR 14.1 for a 100 km reduction and EUR 6.95 for the average gap of 49.3 km. These findings suggest that spatial mismatches impose a meaningful cost on passengers and point to the potential value of multimodal integration in reducing the distance passengers must cover themselves.

Using the estimated parameters, we assess the welfare implications of introducing BlaBlaModes through a series of counterfactual simulations. We progressively reduce the spatial gaps between searched and booked itineraries, from the observed level to their complete elimination, to quantify how improvements in the available itineraries affect consumer choices and welfare. Throughout these simulations, we hold prices and the observed supply of trips fixed, allowing us to isolate the demand-side effects of reducing spatial mismatches. Specifically, the model captures how passengers reallocate their choices across carpooling, buses, and the outside option as the spatial gaps associated with available trips decrease. We measure the resulting changes

in consumer surplus and booking probabilities, and examine how different levels of a potential uniform surcharge applied by the platform affect consumer gains.

The counterfactual analysis shows that eliminating spatial gaps increases consumer surplus by approximately EUR 0.57 per searcher and EUR 5.6 per booking, representing an aggregate gain of EUR 10.4 million across the 18.3 million searches in our sample, equivalent to about a quarter of the average fare per booking and achieved without adding any new transport capacity. These gains reflect the reduction in the disutility associated with imperfect trips. The technology also expands the market by increasing the share of searchers who book a trip from 10.75% to 11.6%. This increase is larger for carpooling than for buses, suggesting that carpooling accounts for the larger share of additional bookings as a direct consequence of reducing spatial gaps.

Lastly, we assess whether the platform could charge for the new functionality while preserving the consumer-surplus gains it creates. We find that a uniform surcharge of approximately EUR 5.12 per booked trip, about the 25% of the average fare, would eliminate the average consumer-surplus gain relative to the observed market. Below this level, average consumer surplus remains higher than under the current technology. The results therefore suggest that there may be scope to charge for multimodal integration while preserving positive average consumer-surplus gains, although the benefits of the technology and the effects of a surcharge may differ across consumers.

Contribution to the literature. This paper contributes to three strands of literature: search and matching frictions in transportation markets, digital transportation platforms and multimodal services, and the environmental implications of shared transport.

First, we contribute to the literature on search and matching frictions in transportation markets. Existing studies have investigated how pricing, entry, and platform design affect matching and market outcomes in taxi and ride-hailing markets. This literature includes work on taxi fares and entry restrictions (Lagos, 2000; Frechette et al., 2019), the structure of taxi fares and search inefficiencies (Buchholz, 2022), competition between traditional taxis and ride-hailing platforms (Bian, 2018; Shapiro, 2018), and pricing and matching on ride-hailing platforms (Castillo, 2025; Gaineddenova, 2022). While these studies provide important insights into the sources of frictions in short-distance transportation markets, less is known about the welfare costs arising when passengers cannot find a trip that matches their searched origin and destination. We focus on a different source of friction: passengers may be unable to find a suitable trip because the searched origin-destination pair is not served or because no suitable option is available at the time of the search. Therefore, passengers may leave the platform without booking or accept a trip whose origins or destinations differ from those they initially searched, incurring an addi-

tional spatial cost. While we do not model the matching process itself, we quantify the disutility associated with these supply-related frictions and study their implications for consumer choices and welfare. We then use a structural demand model to assess the potential gains from multimodal integration, which can expand the set of feasible itineraries by combining services that would not provide a suitable connection on their own. In doing so, we connect the literature on transportation-market frictions to the understudied question of whether complementarity across modes can improve consumer welfare, by reducing these sources of frictions.

Second, we contribute to the literature on digital transportation platforms, with a particular focus on long-distance carpooling and the provision of multiple transport modes within the same platform. Much of the empirical literature on digital transportation platforms focuses on urban and short-distance ride-hailing services, including passenger preferences (Buchholz et al., 2020; Christensen et al., 2023), interactions with public transportation (Hall et al., 2018), congestion (Tarduno, 2021), driver flexibility (Chen et al., 2019), consumer benefits (Cohen et al., 2016), and pricing and peak-demand management (Castillo et al., 2025). The literature on medium- and long-distance carpooling is more recent and has relied primarily on surveys (Delhomme and Gheorghiu, 2016; Gheorghiu and Delhomme, 2018; Monchambert, 2020) or scraped data (Farajallah et al., 2019; Yeung and Zhu, 2022). For instance, Astier et al. (2023) use revealed-preference data from BlaBlaCar to estimate passengers' preferences for ride characteristics. We build on this literature by developing a structural model of transportation-mode choice that accounts for competition between carpooling and bus services on the same platform. The two modes can act as substitutes when passengers choose between alternative services, but also as complements when combining them makes an otherwise unavailable itinerary feasible. This framework allows us to quantify the implications of multimodal integration for consumer choices and welfare.

Third, this paper relates to the literature on the environmental implications of shared transport. Previous studies have assessed the potential of carpooling to reduce emissions and fuel consumption, as well as the role of public policies in promoting shared mobility. Yin et al. (2018), for example, study how carpooling as a mode of transportation can mitigate CO₂ emissions in the Paris region, while Jacobson and King (2009) quantify the fuel savings from increasing vehicle occupancy. We complement this literature by studying whether multimodal integration can make shared transport more attractive to passengers. By allowing carpooling and bus services to complement one another, multimodal technology may enable passengers to complete journeys that would otherwise require a private car or for which suitable shared-transport options are unavailable. Our analysis quantifies consumers' valuation of multimodal integration and pro-

vides evidence on a potential mechanism through which it could incentivize the use of shared transport. We do not directly estimate changes in private-car use or emissions; the environmental implications therefore depend on whether changes in consumers' choices translate into shifts away from private-car travel.

The rest of the paper proceeds as follows. Section 2 describes the platform under study and its characteristics. Section 3 presents the data. Section 4 provides suggestive evidence of heterogeneity and presents key stylised facts behind the willingness to multimode. Section 5 outlines the model of consumer behaviour and discusses identification and the estimation procedure, while Section 6 presents the estimation results. Section 7 details the counterfactual mechanisms and the simulation results, and Section 8 concludes.

2 Background: The BlaBlaCar Platform

We study travel demand on BlaBlaCar, the largest long-distance carpooling platform in the world, operating in more than twenty countries with over one hundred million registered users. Two features of the platform make it a natural setting for studying a multimodal technology. First, it offers two intercity modes on a single interface: peer-to-peer carpooling, in which private drivers offer the spare seats in their car, and scheduled coach services through buses.¹ Second, both sides of the carpooling market are private individuals using the service occasionally. Professional use is prohibited and registration, which is free, is required to post or book. A typical trip is a long-distance journey shared by a driver and passengers who do not know each other, rather than a regular commute.

The platform brings together transportation supply and demand by listing both carpooling and bus services. On the supply side, drivers post upcoming trips with spare seats, specifying the pick-up and drop-off locations, departure time, any intermediate stopovers at which passengers may be picked up or dropped off, the price per seat, the number of available seats, and whether requests are accepted automatically or approved manually. Bus services are listed in a similar way, as scheduled services with an origin, destination, departure time, capacity, and a sequence of intermediate stops. On the demand side, prospective passengers submit a search request by entering a departure date, an origin, and a destination. The platform then returns the carpooling rides and bus services available at the time of the search.²

There are two important features of the searching and booking processes that are relevant of our

¹The bus service is operated under the BlaBlaCar brand following the platform's acquisition of a coach operator; for our purposes what matters is that a passenger searching a route is offered, on the same results page, both a set of carpooling rides and a set of bus departures.

²For a detail description of the matching process see Appendix A.

empirical strategy. First, the ranking of results is mechanical: available rides for the requested day, including both carpooling and bus services, are ordered chronologically by departure time. The position of a ride therefore conveys no information about its predicted relevance or quality. The displayed choice set provides an account of the options available to the passenger at the time of the search. Second, the results page provides a rich set of attributes for each ride, including its price, departure and arrival times, the proximity of the proposed meeting points to the locations entered by the passenger, the driver’s rating and identity, whether booking is automatic or requires approval, and whether the driver commits to leaving the middle back seat empty. By clicking on a ride, the passenger can also see the exact pick-up and drop-off addresses and whether the requested cities are the driver’s own endpoints or stopovers along the route. To book a trip, the passenger sends a booking request; once the driver accepts it, payment is made and contact details are exchanged. After the trip, both parties can rate their experience. The same information is provided for available bus services, and the booking process follows the same steps except that bus bookings are confirmed automatically.

This structure makes two frictions directly observable and motivates the multimodal technology we study. A passenger may search a route for which no suitable single-mode ride exists (an underserved route), or for which rides exist but none matches the exact origin and destination on the desired day (a congested route). This is because of the different timing drivers can post an available trip and that passengers can search for a trip. While the supply of buses is fixed and defined beforehand by the platform, and its availability depends on how many passengers have booked a trip at the time an individual is making the search, drivers can publish a trip anytime before the day before the trip. The heterogeneous entry decision of drivers, and the search by passengers that can happen at any period in time, makes the choice set available to passengers highly dynamic in time, and depending on the time the search is made. Besides, both available buses and cars are dependent on the realized bookings at the time of the search. In either case, the passenger either leaves the platform or books an imperfect substitute displaced in space from the searched trip. Booking a trip that does not perfectly match the searched origin-destination (OD) requires the passenger to reach the chosen pick-up and drop-off stops by covering themselves the distance between the searched and booked origins and destinations. Two consequences are related to this choice. First, the passenger experiences a disutility from the distance they still need to cover, assuming they start their trip from the searched origin and end it at the searched destination. Second, this creates scope for potential gains from multimodal technologies.

The frictions described above motivate our analysis of BlaBlaModes, a multimodal function-

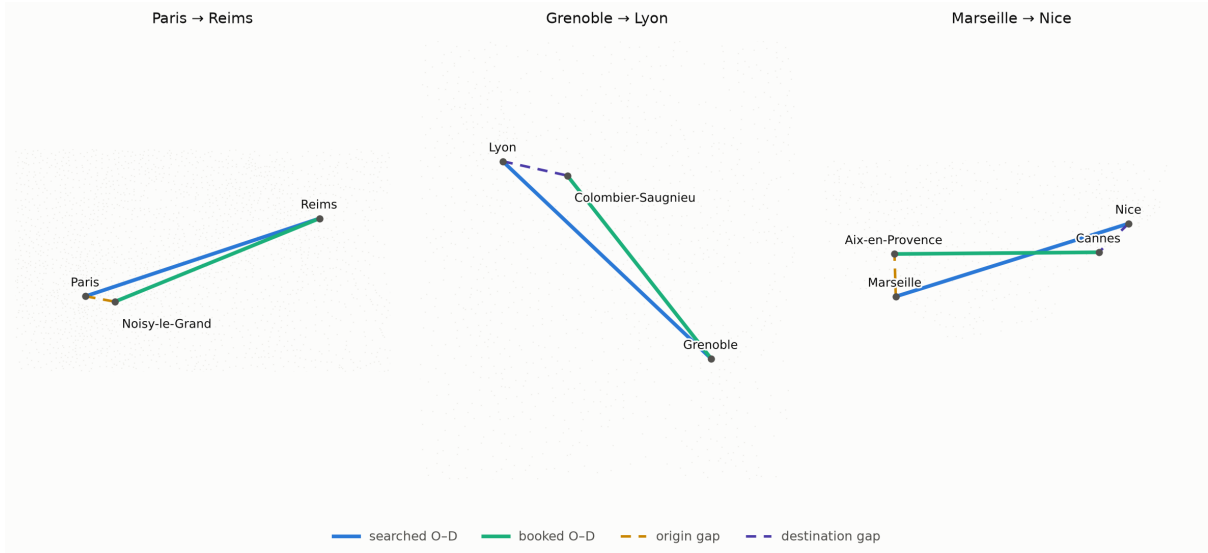


Figure 1
Examples of near-match bookings

Note: Each panel shows an illustrative near-match booking drawn from the data. The blue line is the origin–destination (OD) pair the passenger searched; the green line is the OD actually booked. The orange dashed segment is the origin gap and the purple dashed segment is the destination gap, drawn only when the booked endpoint differs from the searched one. The three panels illustrate the possible cases: an origin gap only, a destination gap only, and gaps at both endpoints. In each case the passenger must cover the dashed distances between the searched and booked endpoints on their own.

ality currently under development at BlaBlaCar. Developed in the context of initiatives to promote more sustainable travel ([Agence de l’Environnement et de la Maîtrise de l’Énergie \(ADEME\), 2022](#)), BlaBlaModes is intended to integrate multimodal itineraries directly into the platform’s search and booking process, allowing passengers to access combined itineraries alongside direct trips. The planned functionality would combine carpooling, buses (including BlaBlaCar Bus), and trains for medium- and long-distance journeys. By connecting services across modes, it could both reduce the spatial gaps passengers need to cover themselves and make additional origin-destination pairs reachable. Although BlaBlaModes has not yet been implemented, it provides a concrete motivation for studying how multimodal integration could affect consumer choices and welfare, and how much consumers might be willing to pay for the resulting improvements in available itineraries.

To illustrate how spatial mismatches arise in the data, Figure 1 shows examples of searched and booked trips. The blue line represents the trip requested by the passenger, connecting the searched origin and destination, while the green line represents the trip actually booked. The dashed lines indicate the distances between the searched and booked origins and destinations that passengers must cover themselves when the corresponding locations do not coincide.

Multimodal integration can expand the set of itineraries available to passengers through two related channels. First, combining different modes can make journeys feasible between origin-

destination pairs that no single mode serves directly. A passenger may, for example, use one mode to reach an intermediate location and another to complete the journey. Second, multimodal integration can expand the choice set on routes that are already served by a single mode. Combining cars, buses, and trains can create additional itineraries with different departure times or routes, increasing the likelihood that passengers find an option suited to their travel needs. Although these itineraries may require connections and involve additional travel time, they can provide alternatives when no suitable direct trip is available at the time of search. The potential value of multimodal integration therefore extends beyond increasing the spatial coverage of the transportation network. By combining existing services, it can both connect previously unserved origin-destination pairs and expand the alternatives available on routes that are already served.

3 Data and Descriptive Statistics

We use proprietary data provided by BlaBlaCar covering the full population of searches, bookings, and related carpooling and bus supply on the 100 most popular routes in France from November 2021 to November 2022. Each observation in the search data corresponds to a passenger search, which may or may not result in a booking. When a booking is finalized, the search is assigned a booking ID that allows us to link the passenger’s initial search to the booking and its associated characteristics. The data therefore provide detailed information on both the demand and supply sides of the market.

The main demand file records individual search requests, with one row per query. It does not contain a passenger identifier, so we treat each search as an independent choice observation. Each record reports the searched origin and destination, the departure date, and the number of days between the search and the departure date. We also observe the choice set displayed by the platform, measured by the number of carpooling and bus results available at the time of the search, along with their average displayed prices. Finally, we observe the number of results clicked and a set of demographic characteristics of potential passengers, specifically gender and age group based on year of birth.

The search data also indicate whether a search resulted in a booking. Using the booking ID, we can link each successful search to the characteristics of the resulting booking. We distinguish between bookings made through the carpooling service and those made through the platform’s bus service. A carpooling booking file contains one row per booking and records the booked origin and destination, the driver’s actual pick-up and drop-off cities, the price components paid by the passenger (the amount paid to the driver, the platform commission, and VAT),

and the number of seats booked and offered by the driver. It also includes a smart-stopover indicator, which identifies cases in which the driver deviated from the originally offered trip to serve the passenger at intermediate stops, as well as cancellation indicators. Finally, the file records characteristics of the driver’s full trip, including its distance and duration, allowing us to locate the booked segment within the driver’s journey.

We use a separate booking-level file for bus services, with one row per bus booking. This file records the origin and destination of the booked bus journey, the number of tickets sold, and the number of days between booking and departure. It also contains the price paid by the passenger, including VAT, the distance and duration of the trip, and an indicator for whether the ticket is a round trip.

Together, these three demand files allow us to reconstruct the booking process associated with each search. We observe what the passenger searched for, the choice set displayed at the time of the search, the resulting booking, and whether the booked origin and destination match those requested. We classify a booking as a near-match when its booked origin or destination differs from the searched one. We also use the total number of searches for each origin-destination and departure-date pair to measure demand for that market, conditional on passengers entering the platform.

On the supply side, we use detailed bus supply data to construct the network of bus services offered by the platform across the 100 main routes. For each origin-destination pair and departure date, we observe the number of bus services covering the corresponding leg, their duration and distance, and their offered capacity. We also observe the number of legs, defined by the origin-destination segments, that make up the main route, as well as the number and ordering of intermediate stops along that route. For each origin and destination, we can additionally compute the number of buses and cars departing from or arriving at that location. These measures allow us to characterize the role of each location in the network, in a way analogous to the hub-and-spoke structure of the airline market. The bus inventory measures the supply of scheduled services on a route independently of realized bookings and is analogous to the information on driver-posted trips available for carpooling.

For carpooling, however, we observe only the total number of cars offered for a given origin-destination pair and departure date. We do not observe the characteristics of drivers whose trips were offered but not booked. Driver and trip characteristics are available only for trips associated with bookings.

Combining the demand and supply data allows us to construct two measures of supply: one based on the inventory of offered services and another based on the options displayed to passen-

gers at the time of their search. For each market and mode, we use the supply data to measure the number of carpooling trips and bus services offered on the searched route. We then use the search data to measure the number of carpooling and bus results displayed to passengers. The latter measure captures the availability of options at the time of the search and can vary with the timing of supply entry and realized bookings.

Comparing these measures helps distinguish between two types of market frictions relevant to multimodal integration. The inventory indicates whether a route is served by a given mode, while the number of displayed results indicates whether options are available to passengers when they search. If no results are displayed for a mode but its inventory is positive, the route is served, but no options are available at the time of the search. We interpret this as a congestion-related friction. If the inventory is also zero, the searched origin-destination pair is not served by that mode. These measures therefore allow us to distinguish between the absence of supply and the lack of available options at the time of search.

We also construct a measure of the spatial distance between searched and booked origins and destinations. This measure captures the extent to which a passenger’s booked itinerary differs from the requested one and allows us to quantify the spatial gaps associated with near-match bookings. We assign each searched and booked location to a French municipality (*commune*) using official geographic data.³ For each booking, we compute the distance between the searched and booked origins (the origin gap) and between the searched and booked destinations (the destination gap). If the searched and booked endpoints belong to the same municipality, the corresponding gap is zero. Positive gaps measure the spatial mismatch between the requested itinerary and the substitute trip accepted by the passenger.

We augment the data with information on daily fuel prices. These data come from the French government’s open-data portal, which records price changes reported by fuel stations across France. We construct a daily panel of diesel and gasoline (E10) prices from November 2021 to November 2022.⁴ For each municipality, we calculate the daily fuel price as the average across stations within 20 km or, if no station lies within that radius, across the three nearest stations. We then match these prices to the origin and departure date of each trip.

For identification, we restrict the sample to searched origin-destination pairs with at least one observed booking during the sample period. We also exclude a small set of bookings that are implausible as mode substitutions. Specifically, we drop a booking only when both its origin and destination differ from those searched, at least one mode had supply available on the searched

³Municipality coordinates come from La Poste’s *Base officielle des codes postaux*, available at data.gouv.fr.

⁴The fuel price data are available at prix-carburants.gouv.fr. Since stations report a price only when it changes, we carry each station’s last reported price forward for up to 21 days when no new price is reported.

Table 1

Descriptive statistics: searches, mode choice, and demographics

	Mean	Std. Dev.
Number of searches	18,259,760	
Number of bookings	1,163,897	
Bus (%)	3.4	18.2
Carpool (%)	2.9	16.9
Outside option (%)	93.6	24.4
Women (%)	49.1	50.0
Age 10–19 (%)	9.2	28.9
Age 20–29 (%)	39.3	48.8
Age 30–39 (%)	12.3	32.8
Age 40–49 (%)	5.8	23.4
Age 50–59 (%)	3.3	17.9
Age 60+ (%)	1.3	11.5
Age unreported (%)	28.9	45.3

Notes: The number of searches is the total number of searches for the origin-destination routes and departure dates analyzed (the sum of market-level search counts, where a market is a searched route by departure date). Bus, carpool, and outside-option shares are calculated relative to searches; the outside option is the share of searches that did not result in a booking. Demographics are measured in the booking data. The share of women is calculated among bookings with reported gender, and age is measured in 10-year groups, with ages 60 and above pooled. For a share p , the reported standard deviation is that of the underlying binary indicator, $\sqrt{p(1-p)}$. Carpool refers to BlaBlaCar’s ride-sharing service.

route at the time of the search, and the sum of the origin and destination gaps exceeds 200 km. On a route where supply was both offered and available, such a booking is more plausibly an unrelated trip than a supply-driven mismatch. We retain bookings with similarly large gaps when no supply was available on the searched route, since these may represent substitutions in unserved markets.

Table 1 reports descriptive statistics for the final sample. The sample contains more than 18 million searches and approximately 1.16 million bookings. The overall conversion rate is low: about 6.4% of searches result in a booking. This rate should be interpreted in light of the nature of platform searches, which include exploratory queries and searches that do not necessarily reflect a firm intention to travel. Bookings are divided relatively evenly between buses and carpooling. Among passengers with reported demographic information, men and women are both represented, while bookings are concentrated among younger age groups. This pattern is consistent with the use of long-distance carpooling and bus services by younger travelers.

Table 2 reports descriptive statistics for the route-day-time cells covered by BlaBlaCar services in our sample. The top panel reports the prices paid by passengers. Average bus prices are lower than average carpooling prices. For carpooling, the table also reports the decomposition of the total price into the amount paid to the driver, the platform commission, and VAT.

Table 2

Descriptive statistics: route–day–time of day level

	Mean	Std. Dev.
Number of observations	79,970	
Price bus (EUR)	20.34	21.77
Price carpool (EUR)	24.39	16.51
Driver fee (EUR)	18.13	12.34
Commission (EUR)	5.32	3.57
VAT (EUR)	0.93	0.66
Monday (%)	14.8	35.5
Tuesday (%)	11.5	31.9
Wednesday (%)	11.8	32.3
Thursday (%)	14.1	34.8
Friday (%)	16.6	37.2
Saturday (%)	15.4	36.0
Sunday (%)	15.8	36.5
Departure 0h–6h (%)	17.1	37.6
Departure 6h–12h (%)	31.5	46.4
Departure 12h–18h (%)	30.8	46.2
Departure 18h–24h (%)	20.7	40.5
Trip distance (km)	273.7	176.4
Trip duration (hours)	3.13	1.87
Total distance gap (km)	49.3	92.9

Notes: The unit of observation is a searched origin-destination route by departure date by time-of-day block (six-hour blocks). Bookings are collapsed to this level by averaging within each cell before computing the statistics. Prices are the total prices paid by consumers. The carpool price is decomposed into the driver fee, the platform commission, and VAT charged on the commission; bus tickets carry no commission or VAT. Bookings with missing price components are excluded from the price rows. Day-of-week and time-of-day rows report the share of observations in each category; for a share p , the reported standard deviation is that of the underlying binary indicator, $\sqrt{p(1-p)}$. Trip distance and duration are platform-recorded. The total distance gap is the sum of the origin and destination great-circle distances between searched and booked endpoints and is zero for perfectly matched bookings. Carpool refers to BlaBlaCar’s ride-sharing service.

The middle panel shows how observations are distributed across days of the week, while the next panel reports their distribution across six-hour departure-time blocks. The distribution is broadly spread across days and time periods, with departures somewhat more concentrated in the morning and afternoon. Finally, booked trips are relatively long in both distance and duration, consistent with the platform’s focus on long-distance travel. The average total spatial gap between searched and booked endpoints is 49.3 km.

4 Preliminary Evidence

In this section, we first document the magnitude of origin-destination near-matches and the potential scope for multimodal technologies to improve the fit between passengers’ desired and

available itineraries. We then examine how near-matches vary across origin-destination pairs, transport modes, and the distance between searched and booked endpoints. This heterogeneity motivates the variation we exploit in our estimation. We next present suggestive evidence on the determinants of near-matched listings and bookings, focusing on supply frictions. Finally, we examine passengers’ willingness to accept alternatives when an exact match is not available.⁵ Throughout, a *near-match* denotes a booking whose origin-destination pair differs from the one the passenger searched.

4.1 Magnitude and Heterogeneity in Origin-Destination Mismatches

We begin by documenting the prevalence of near-matches and how they vary across bookings. These patterns provide descriptive evidence on the extent of spatial mismatches and motivate the empirical strategy.

Near-matches are not an isolated phenomenon but an important feature of the market. Table 3 reports their distribution across bookings and transport modes. Overall, 34% of bookings involve an origin or destination that differs from the one initially searched. Near-matches also arise in the choice sets displayed to passengers whose searches do not result in a booking: among the alternatives displayed, 13% of carpooling options and 16.8% of bus options are near-matches. These figures highlight the gap between the itineraries passengers initially request and the alternatives available on the platform.

The prevalence of near-matches differs substantially across modes. Only 17.8% of bus bookings are near-matches, compared with 53.1% of carpooling bookings. This difference is consistent with the greater spatial flexibility of carpooling. Drivers choose their own routes and may accommodate passengers through intermediate stopovers, whereas bus services operate on fixed routes with scheduled stops. Carpooling therefore offers more opportunities to adjust pick-up and drop-off locations when the exact searched itinerary is unavailable. Nevertheless, the high share of near-matches in both modes indicates that passengers often accept trips that do not perfectly match their requested origin and destination.

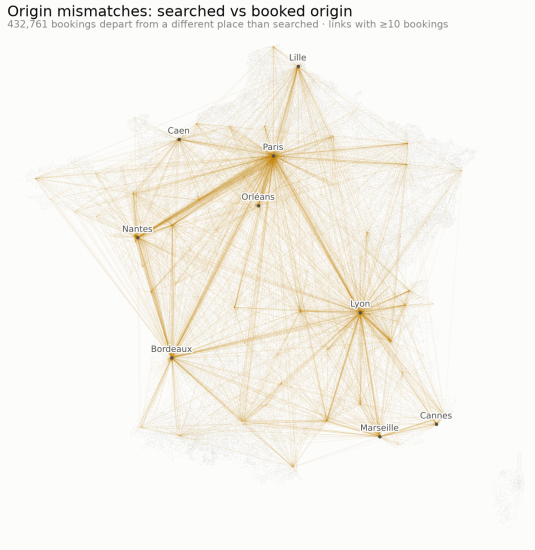
Near-matches also vary substantially across locations. Figure 2 maps near-matched bookings involving a different origin (Panel a) and a different destination (Panel b). The size of each node represents the number of near-matched bookings associated with that location. Near-matches are particularly frequent around large cities and in southern France, but they are not confined to a small number of routes. Instead, they occur throughout the French transportation network, with substantial variation in their incidence across origins and destinations. This geographical

⁵All statistics reported in this section are based on the full sample of searches and bookings.

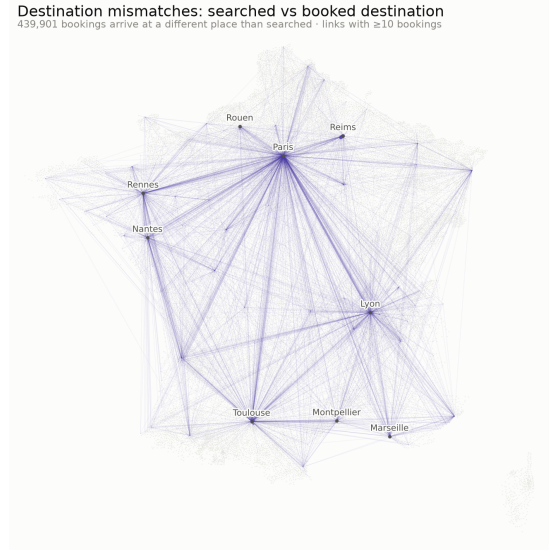
Table 3
Search–booking gaps in booked trips

	Total	Bus	Carpool
Number of bookings	1,163,897	629,626	534,271
Search–booking match (%)	66.0	82.2	46.9
Origin gap only (%)	15.1	5.8	26.1
Destination gap only (%)	16.1	11.9	21.0
Origin and destination gaps (%)	2.9	0.2	6.1
Trip distance (km)	210.0	223.0	200.0
Origin gap (km)	12.5	133.7	10.2
Destination gap (km)	10.7	5.8	13.0
Origin + destination gap (km)	46.2	23.2	46.9

Notes: A booking has an origin (destination) gap when the booked origin (destination) differs from the searched one. The shares of exact matches, origin-only gaps, destination-only gaps, and gaps at both endpoints sum to approximately 100%, allowing for rounding. Distance rows report medians. Trip distance is the platform-recorded trip length. Origin and destination gap distances are great-circle distances between the searched and booked endpoints, computed among bookings with the corresponding gap. The last row reports the sum of the origin and destination gaps among bookings with gaps at both endpoints. Carpool refers to BlaBlaCar’s ride-sharing service.



(a) Origin mismatches



(b) Destination mismatches

Figure 2

Geographical distribution of origin and destination mismatches

Note: Each panel maps the geographic distribution of near-match bookings across France. In Panel (a), a booking is counted at its searched *origin* whenever the booked origin differs from the one searched; in Panel (b), it is counted at its searched *destination* whenever the booked destination differs. The size of each node is proportional to the number of near-match bookings associated with that location. Based on the full sample of bookings.

heterogeneity motivates our empirical analysis of how the availability of near-match alternatives varies across markets.

The spatial extent of near-matches also varies considerably. Among bookings with a gap only at the origin, the median distance between the searched and booked origins is 12.5 km. Among

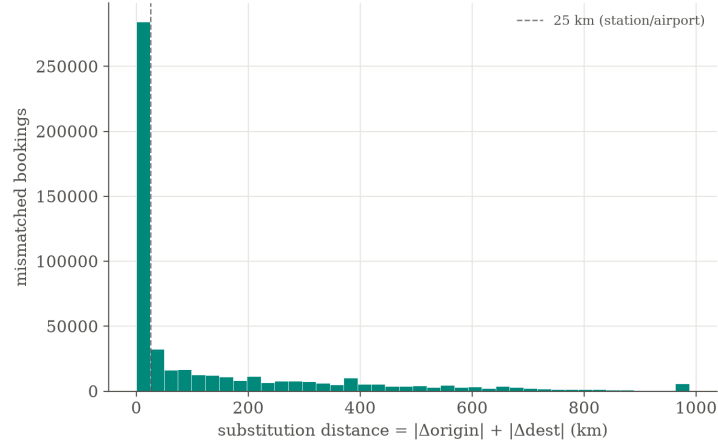


Figure 3
Distribution of spatial gaps in near-matched bookings

Note: The figure reports the cumulative distribution of the sum of the origin and destination gaps between searched and booked endpoints.

bookings with a gap only at the destination, the corresponding median destination gap is 10.7 km. For bookings with gaps at both endpoints, the median sum of the two gaps is 46.2 km. These figures also reveal important differences across modes. For bus bookings with an origin gap, the median origin gap is much larger, at 133.7 km, consistent with passengers using a different location, potentially a transport hub, instead of the searched origin. For carpooling, the median origin gap is only 10.2 km, consistent with drivers' greater flexibility in serving nearby locations.

The distribution of total spatial gaps further highlights the heterogeneity of these substitutions. Across near-matched bookings, the mean sum of the origin and destination gaps is approximately 139 km, compared with a median of about 18 km. Roughly 55% of near-matches involve substitutions within 25 km, while the remaining 45% involve larger deviations. Figure 3 plots the cumulative distribution of near-matched bookings by the sum of the distances between searched and booked endpoints. The distribution is strongly right-skewed: most substitutions involve relatively short distances, but a non-negligible share involve much larger deviations from the originally searched itinerary.

The distribution suggests that near-matches capture different types of passenger adjustments. Some involve relatively small geographic substitutions, which may reflect the spatial organization of transport infrastructure rather than a substantial change in the passenger's intended destination. Examples include booking a nearby station, airport, or suburban terminal instead of the location originally searched, such as Bezannes instead of Reims, Lyon Saint-Exupéry instead of Lyon, or Orly instead of central Paris. Other near-matches involve substantially larger deviations from the searched itinerary and may reflect more significant adjustments to travel

plans when the preferred itinerary is unavailable.

Overall, the descriptive evidence shows that near-matches are both prevalent and highly heterogeneous. Passengers frequently accept spatially imperfect substitutes, but the extent of the adjustment varies considerably across modes, locations, and bookings. This variation motivates a framework in which the value of multimodal technology depends not only on whether an itinerary is available, but also on how closely the available alternative matches the passenger’s desired origin and destination.

4.2 Determinants of Frictions

We now examine descriptive evidence on the mechanisms that may drive near-matches, focusing on supply-side frictions.

Table 4 decomposes near-matches using two measures of supply: available supply at the time of the search and total supply. We distinguish three categories: (i) *underserved*, when the searched OD has no offered supply of the booked mode; (ii) *congestion*, when supply exists but no options are displayed; and (iii) *other*, when options are displayed for the searched OD but the traveller books a different OD. The reasons for near-matches differ substantially across modes. For buses, almost half of near-matches (50%) occur because the searched OD has no bus supply. In these cases, passengers may use a nearby hub instead, consistent with the hub-and-spoke structure of the bus network. For carpooling, by contrast, only 9% of near-matches are due to the lack of supply on the searched OD, while 91% occur even though carpooling options were available. Congestion is relatively rare for both modes, accounting for less than 3% of near-matches.

These differences suggest that near-matches arise for different reasons across modes. For buses, a substantial fraction reflects the spatial coverage of the network: passengers whose searched locations are not directly served may use a nearby hub. For carpooling, most near-matches occur even when an option is available on the searched OD. This suggests that the timing and characteristics of available trips, rather than coverage alone, may play an important role in the choice of an alternative OD. This distinction matters for the potential role of multimodal integration. Combining the two modes could expand the set of feasible trips by using the spatial coverage of one mode to complement the other, but it could also provide alternatives when a direct option does not match the traveller’s desired itinerary.

We find a similar pattern among searches that do not result in a booking. Table 5 applies the same supply measures to all searches, distinguishing between searches for which the OD is not served by a mode, searches for which the OD is served but no option is available at the time of the search, and searches for which options are displayed but the passenger does not book.

Table 4

Reasons for booking a mismatch (estimation sample)

Service	Near-matched bookings	Underserved (no supply)	Congestion (supply, no options)	Other (options existed)
Carpooling	283,901	24,338 (8.6%)	2,438 (0.9%)	257,125 (90.6%)
Bus	111,913	55,663 (49.7%)	2,483 (2.2%)	53,767 (48.0%)

Denominator: bookings whose booked OD differs from the searched OD.

Underserved: the searched OD has no offered supply of the booked mode.*Congestion*: supply exists on the searched OD, but no options were displayed at the time of the search.*Other*: options were displayed for the searched OD, but the user booked a different OD.

Table 5

Reasons for not booking (all searches)

Service	Searches	Underserved	Congestion	Other
Carpooling	20,621,323	1,666,190 (8.1%)	2,273,139 (11.0%)	16,681,994 (80.9%)
Bus	20,621,323	5,080,873 (24.6%)	1,086,313 (5.3%)	14,454,137 (70.1%)

Percentages are calculated relative to all searches for the corresponding mode.

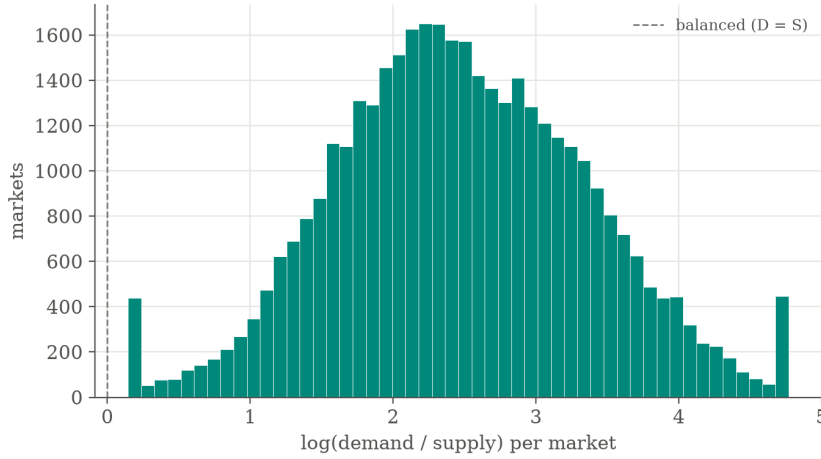
Underserved: the searched OD is not served by this mode.*Congestion*: the searched OD is served, but no options were available at the time of the search.*Other*: options were displayed, but the user did not book.

In both modes, the last category is the largest: 81% of carpooling searches and 70% of bus searches end without a booking despite visible options. At the same time, lack of coverage is more important for buses: about 25% of bus searches concern an OD that is not served, compared with only 8% of carpooling searches. Taken together with Table 4, this evidence suggests that the existence of a connection is only part of the choice problem. Even when a connection exists, its timing and routing may not provide a sufficiently good match to the traveller’s desired trip.

Table 5 also highlights the role of supply timing. For carpooling, 11% of searches do not result in a booking because no carpooling option is available at the time of the search, compared with 5% for buses. Thus, even when an OD is served by a mode, the options available to a passenger depend on when the search is made. More than 15% of searches concern an OD that either has no supply or has no available option at the time of the search. This points to the importance of dynamic choice sets: the relevant supply for a passenger is not the set of all trips offered on a route, but the subset available when the passenger searches. Multimodal technology could affect this margin by combining services across modes and thereby increasing the set of itineraries available to a passenger at a given time.

Figure 4 provides further evidence on the relationship between search demand and displayed supply. The distribution of $\log(\text{demand}/\text{supply})$ lies almost entirely above zero, indicating that search demand exceeds displayed supply in nearly all markets, according to these measures. This

Figure 4
Supply-demand imbalance across markets



Note: The figure plots the distribution across markets of $\log(\text{demand}/\text{supply})$, where a market is a searched route by departure date. Demand is the number of searches in the market and supply is the number of displayed results (carpooling plus bus) available to searchers in that market. Values above zero indicate markets in which search demand exceeds the supply displayed to passengers; the distribution lies almost entirely above zero, so displayed options fall short of search demand in nearly all markets. Because supply is measured as the options displayed at the time of the search, rather than the total number of trips offered on the route, the ratio reflects the availability passengers actually face. Based on the full sample of searches.

pattern is consistent with displayed choice sets being small relative to the number of searches in a market. It also reinforces the importance of supply timing: the number of trips offered on a route does not necessarily correspond to the number of options displayed when a traveller searches.

The difference between offered and displayed supply provides another indication of the importance of dynamic availability. Figure 5 compares the number of options displayed to passengers with total supply on the same trip. For buses, total offered supply is determined in advance by the firm when setting its schedule. For carpooling, by contrast, offered supply is determined by drivers and observed ex post. In both cases, however, the number of options displayed to a passenger at the time of the search can differ from total offered supply. Some bus seats may already have been booked, while some carpooling trips may not yet have been posted by drivers. The set of options faced by a passenger therefore depends not only on total supply on the route, but also on when the search takes place relative to the departure date.

The distributions of the displayed-to-offered supply ratio differ across modes. For buses, the median ratio is about 0.56 (Figure 5, left), meaning that the number of displayed options is around 56% of the measured offered supply at the median. For carpooling, the median ratio is above one, at about 1.37 (Figure 5, right). This difference partly reflects the different definitions of offered supply across the two modes, so the ratios should not be interpreted as a direct comparison of the amount of supply available to passengers. In both distributions, a non-

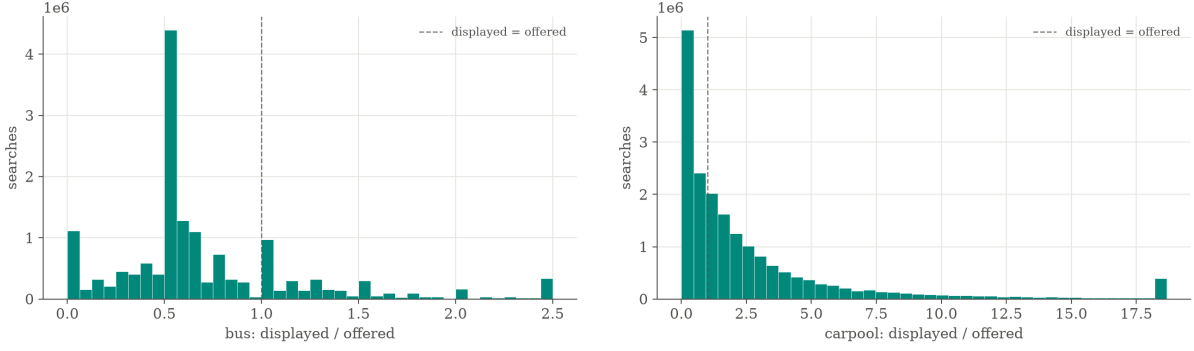


Figure 5
Displayed relative to offered supply, by mode

Note: Each panel plots the distribution across searches of the ratio of displayed to offered supply of a mode: bus (left) and carpooling (right). Displayed supply is the number of results of the mode shown to the traveller at the time of the search; offered supply is the total number of services of that mode on the searched route (the scheduled services for buses and the posted trips for carpooling). The median ratio is about 0.56 for buses and about 1.37 for carpooling. Because offered supply is measured differently across the two modes, the ratios should not be compared directly across panels; in particular, a carpooling ratio above one indicates that its numerator and denominator do not measure the same object. In both panels a non-negligible mass sits at zero, corresponding to searches for which no option is displayed even though offered supply is positive. Based on the full sample of searches.

negligible share of searches has a ratio equal to zero, indicating that no options are displayed even though the supply measure is positive.

Overall, this descriptive evidence points to an important role for the temporal availability and coordination of supply. Near-matches do not arise only because a searched OD is not served: in many cases, options are displayed for the searched OD, but passengers book a different itinerary. At the same time, the options displayed to passengers vary over time and may differ substantially from total offered supply. The relevant friction is therefore not simply whether a connection exists, but whether an itinerary matching the traveller’s desired origin, destination, and timing is available when the search takes place. Multimodal technology could relax this constraint by combining supply across modes and increasing the set of itineraries available to passengers.

4.3 Willingness to multimode

The evidence above shows that passengers often book an itinerary that does not exactly match the OD they initially searched. For multimodal technology to improve consumer outcomes, however, passengers must be willing to consider alternative itineraries when a suitable direct connection is unavailable. To motivate our structural model, we therefore examine how booking choices vary with the connectivity of the searched endpoints to other origins and destinations. Specifically, we estimate probit regressions relating endpoint connectivity, through either cars or buses, to (i) the probability of booking and (ii) the probability of booking a near-match.

Our main question is whether the likelihood of accepting a near-match depends on how well the searched endpoints are connected to the rest of the transportation network. The idea is that passengers may be more willing to accept a different OD when the searched origin and destination provide access to other locations, making it easier to cover the gap between the searched and booked trips.

For each searched origin–destination (OD) pair and departure date, we measure endpoint connectivity as the number of cars and buses that connect to the searched origin and destination on the departure date, excluding the searched route itself (leave-one-out). Better-connected endpoints provide more opportunities to reach other locations and may therefore make a near-match more attractive. We test whether higher endpoint connectivity is associated with a higher probability of booking and, conditional on booking, with a higher probability of accepting a near-match.

We estimate the following probability model at the passenger–OD level:

$$y_{i,OD} = \begin{cases} 1 & \text{with probability } P(Y = 1 \mid \text{Connectivity}_{o,d}, X, v), \\ 0 & \text{with probability } 1 - P(Y = 1 \mid \text{Connectivity}_{o,d}, X, v). \end{cases} \quad (1)$$

where

$$\begin{aligned} P(Y = 1 \mid \text{Connectivity}_{o,d}, X, v) = G(&\beta_1 \text{Connectivity}_{o,d} \\ &+ \beta_2 \text{Connectivity}_{o,d} \times \text{Unserved}_{o,d} \\ &+ \beta_3 X' + v_{i,od}), \end{aligned} \quad (2)$$

and $G(\cdot)$ is the standard normal CDF in the probit model (our main specification) and the logistic CDF in the logit model. Here, $\text{Connectivity}_{o,d}$ measures connectivity at the searched endpoints, defined as the sum of departing and arriving cars and buses; $\text{Unserved}_{o,d}$ indicates whether the searched OD has no direct supply; and X' includes displayed price and displayed direct supply. Standard errors are clustered at the origin.

We consider two outcomes. The first is conversion, defined over all searches, and indicates whether a search results in a booking. The second is the probability of a near-match conditional on booking, indicating whether the booked OD differs from the searched OD. We also consider a more restrictive outcome that identifies near-matches with a total gap exceeding the median substitution distance. The first outcome captures the relationship between connectivity and the probability of booking at all, while the second captures its relationship with the decision to

Table 6

Endpoint connectivity, conversion, and the propensity to book a near-match

	Probit (main)			Logit		
	Pooled	Bus	Car	Pooled	Bus	Car
<i>Panel A. Conversion: $Pr(book)$, all searches</i>						
All searches	−0.096***	−0.075*	−0.023	−0.182***	−0.138*	−0.049
Carpool	−0.216***	−0.411***	+0.187**	−0.438***	−0.854***	+0.380**
Bus	+0.061	+0.391***	−0.306**	+0.137	+0.865***	−0.679**
<i>Panel B. Any near-match: $Pr(near-match \mid booked)$</i>						
All bookings	+0.131	+0.120	+0.018	+0.223	+0.182	+0.049
Carpool	+0.132***	+0.018	+0.114	+0.210***	+0.028	+0.183
Bus	+0.087	+0.267	−0.131	+0.158	+0.502	−0.244
<i>Panel C. Severe near-match (vs. matched bookings)</i>						
All bookings	+0.131***	+0.085	+0.046	+0.255***	+0.144	+0.110
Carpool	+0.234***	+0.115	+0.116	+0.393***	+0.191	+0.193
Bus	−0.072	−0.198	+0.095	−0.140	−0.377	+0.174

Each cell reports the connectivity coefficient from a separate regression of the panel’s binary outcome. “Pooled” reports the coefficient on total endpoint connectivity; “Bus” and “Car” report the coefficients on its two components in the split specification. Rows indicate the estimation sample (all, carpool, or bus). Probit is the main specification; logit estimates are reported alongside. All regressions include displayed price and log direct supply, and standard errors are clustered at the origin. Standard errors and all other coefficients, including controls, the underserved interaction, and the two linear-probability variants with OD and region fixed effects, are reported in Appendix B. $N \approx 20.6$ million searches (Panel A), 1.30 million bookings (Panel B), and 1.07 million bookings (Panel C). The nonlinear models are estimated on a capped subsample; see Appendix B.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

accept an itinerary that does not exactly match the searched trip.⁶

Table 6 reports the connectivity coefficients, pooled and split by connecting mode, for both estimators. The full estimation tables, including all coefficients, the four specifications, and the two linear-probability variants with origin/destination and region fixed effects, are reported in Appendix B.

On the conversion margin (Panel A), endpoint connectivity is negatively associated with booking in the pooled specification. The mode-specific estimates help explain this result. For carpooling searches, carpool connectivity is positively associated with conversion (+0.19), while bus connectivity is negatively associated with conversion (−0.41). The same pattern appears for bus searches: bus connectivity is positively associated with bus conversion (+0.39), while carpool connectivity is negatively associated with bus conversion (−0.31). Thus, connectivity to the same mode is positively associated with conversion, while connectivity to the other mode is negatively associated with conversion. This pattern is consistent with better connectivity

⁶These regressions are descriptive rather than causal. The location of supply is endogenous, and more connected origins and destinations are likely to differ systematically in demand and other characteristics. We therefore interpret the coefficients as evidence on the relationship between endpoint connectivity and the propensity to accept a near-match, rather than as causal effects. The structural model builds on the same motivation but explicitly models demand for different itineraries and allows us to evaluate the welfare implications of multimodal integration.

increasing the availability of alternatives within a mode while also increasing competition from the other mode.

The results differ when we examine the decision to accept a near-match. Conditional on booking, endpoint connectivity is positively associated with the probability of a near-match for carpooling. The coefficient is positive and statistically significant for both any near-match (+0.13 in the probit and +0.21 in the logit) and severe near-matches (+0.23 and +0.39). For buses, by contrast, the estimates are not statistically distinguishable from zero. This difference is consistent with the evidence in Section 4.2. Most carpooling near-matches occur even when an option is available on the searched OD, whereas about half of bus near-matches are associated with the absence of direct bus supply. Connectivity is therefore more strongly associated with the decision to accept a different itinerary for carpooling, while bus near-matches more often reflect the spatial coverage of the network.

These estimates are consistent with the idea that endpoint connectivity may facilitate substitutions to alternative itineraries. When the endpoints of a searched trip connect to a larger set of locations, passengers who do not find a suitable direct option have more opportunities to combine their booked trip with another service to reach their desired destination. This relationship is particularly evident for carpooling, where near-matches are less directly associated with the absence of supply. The full specifications in Appendix B report a similar sign for the connectivity coefficient in the region-fixed-effects linear-probability model, which controls for distance and market size while preserving variation in endpoint connectivity. Figures 11–12 in the appendix show the corresponding dose–response relationships.

Finally, Figure 6 re-estimates the near-match model separately for each origin $\times$ mode unit with at least fifty bookings. The figure shows substantial heterogeneity across locations. The baseline propensity to book a near-match, reported in the left panel, varies considerably across origin–mode units and is generally higher for carpooling than for buses. The estimated relationship between endpoint connectivity and the probability of a near-match also varies across locations, ranging from negative to positive values, with most estimates at or above zero. A likelihood-ratio test rejects the hypothesis of a common response across units ($p \approx 0$). These results suggest that both the propensity to accept a near-match and its relationship with endpoint connectivity vary across places and modes. This heterogeneity matters for the structural analysis, since the value of integrating the two networks need not be the same across locations.

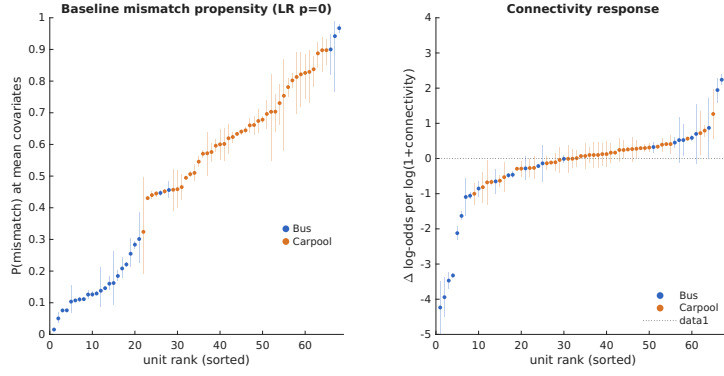


Figure 6
Heterogeneity in the propensity to book a near-match

Note: Each point represents an origin $\times$ mode unit (bus in blue, carpool in orange) with at least fifty bookings, estimated from a separate logit regression of the near-match outcome. Vertical bars are 95% confidence intervals; units are sorted by the plotted estimate. *Left:* baseline near-match probability at mean covariates.

Right: the response of the near-match probability to endpoint connectivity (Δ log-odds per unit of $\log(1 + \text{connectivity})$; dotted line at zero). A likelihood-ratio test rejects a common response across units ($p \approx 0$).

5 Empirical model

We now present our model of consumer choices across transportation modes. Section 5.1 introduces the empirical model, Section 5.2 defines the main specification and discusses identification, and Section 5.3 describes the empirical strategy.

5.1 Demand

We define a market $m = (o, d, t)$ as a long-distance (inter-city) trip from origin o to destination d at departure date t . Consumers wishing to book a trip in market m face a consideration set that comprises trips departing from origin o and arriving at destination d (exact-match trips); alternative trips departing from origin o' and arriving at destination d' , where either (or both) endpoints might not necessarily coincide with the market origin-destination (near-match trips); and the outside option.

On online platforms, a search algorithm determines the consideration set based on trips characteristics, trip availability and the (lack of) supply of trips. Accordingly, there are various reasons why consumers searching in market m may face (and choose) trips whose endpoints differ from the ones they searched for. First, a near-match trip may have a combination of attributes that appeal to consumers enough to compensate for the uncovered leg(s) of the journey. For instance, consumers with strong preferences for departure time and/or transportation mode may choose a trip with those attributes even if it does not start from (or end at) the desired endpoint. Second, exact-match trips may be sold out when consumer makes their decision and such unavailability

increases the probability of having near-match alternatives in the consideration set.⁷ Finally, the requested origin-destination pair may have no exact-match trips at all. In such cases, the search algorithm shows more near-match trips with endpoints farther from the requested ones. The potential trip unavailability or lack of supply imply that consumers face a restriction on their available choices. They are only able to choose among the trips offered when they make their decision, i.e. the number days ahead of the departure date. Therefore, the consideration set is likely to vary over time and across consumers.⁸ Accordingly, we assume that the indirect utility that consumer i derives from booking the trip j in market m is given by

$$U_{ijm} = V(p_{jm}, x_{jm}, G_{jm}, z_i) + e_{ij} \mid j \in J_{im},$$

where p_{jm} denotes price, x_{jm} are trip characteristics (e.g. transportation mode), and z_i represent the observed characteristics of the consumer i . The distance between market origin-destination pair and the trip-specific endpoints is captured by G_{jm} .⁹ Such distance measure is positive for near-match trips and zero for exact-match trips. Notice that consumer i can only consider product j if it belongs to the consumer-specific choice set. We abstract from this aspect in what follows but discuss its implications in Section 7.2. $V(\cdot)$ is a calculable function up to parameters of the observable data and e_{ij} is defined as the difference that makes the equation an identity. We assume that indirect utilities are linear in trip characteristics and specify $V(\cdot)$ and e_{ij} as follows:

$$U_{ijm} = \alpha_i p_{jm} + \beta_i x_{jm} + \gamma G_{jm} + \xi_{jm} + \varepsilon_{ijm}.$$

The error term includes two terms: ξ_{jm} , which captures unobserved product characteristics; and ε_{ijm} , which is the unobserved idiosyncratic taste for trip j . Taste parameters α_i and β_i represent the heterogeneous taste of consumer i and are specified as

$$\alpha_i = \alpha + \pi^p d_i$$

$$\beta_i = \beta + \pi^x d_i,$$

where α is the price sensitivity parameter that is common across consumers, and π^p is a vector of

⁷The current analysis abstracts from dynamics involving the purchasing horizon. See Section 7.2 for a detailed discussion.

⁸This restriction follows the literature on limited choice sets (e.g. Goeree (2008); Hidalgo and Sovinsky (2023)). Unlike previous work, we not only observe the set of trips available for each consumer, but we also observe 1) how the consideration set changes over the purchasing horizon, and 2) how the trips are ranked to the consumers. Exploiting these dimensions is left for future research.

⁹In Section 5.2, we provide details on alternative measures that can capture the origin-destination gap.

parameters that dictate how price sensitivity changes with respect to consumer demographics d_i . The parameter β_i captures similar heterogeneous valuations but related to trip characteristics x_{jm} .

The outside option ($j = 0$) represents a collection of alternatives that range from using own car, choosing other unobserved modes of transportation (e.g. high-speed trains), making the trip with competing inter-city firms, and not doing the journey. We normalize the indirect utility of the outside option and denote it by $U_{i0m} = \varepsilon_{i0m}$. To derive the choice probability, we follow the convention of decomposing U_{ijm} as

$$U_{ijm} = \delta_{jm} + \mu_{ijt} + \varepsilon_{ijm},$$

where $\delta_{jm} = \alpha p_{jm} + \beta x_{jm} + \gamma G_{jm} + \xi_{jm}$ is the alternative-specific fixed effects for each trip in each market. It captures the valuation of all characteristics (observed and unobserved) of trip j that are common to all consumers in market m . The net heterogeneous component is represented by $\mu_{ijt} = \pi^p d_i p_{jm} + \pi^x d_i x_{jm}$ and allows consumer demographics to affect preferences for price and characteristics.

Let consumer i in market m choose trip j with probability

$$s_{ijm} = \int_{A_{ijm}} dF(\varepsilon_i),$$

where $A_{ijm} = (\varepsilon_i | U_{ijm}) > U_{ikm} \forall k \neq j$ and $dF(\varepsilon_i)$ is the density of the unobserved idiosyncratic taste, which is uncorrelated with price. Following Goolsbee and Petrin (2004), we do not impose restrictions on unobserved taste heterogeneity and estimate a more flexible multivariate normal variance-covariance matrix. In doing so, we allow unobserved preference heterogeneity to vary trip by trip and to be correlated across them. For example, a strong preference for early trips by bus may imply strong preference for early trips by carpool services.

Since the choice data can only identify relative rankings, we normalize ε_i with respect to the outside option. The resulting normalized error is distributed multivariate normal with a covariance matrix containing parameters σ .

5.2 Specification and Identification

Specification. The demand model follows the utility specification in Section 5.1. The vector of trip characteristics x_{jm} contains a dummy for the transportation mode, inter-city bus, with carpooling as the base category, and the distance of the searched origin–destination route. We include month, weekday, and departure-window fixed effects and, in our preferred specification,

origin–destination (route) fixed effects, which absorb route distance. Departure time enters through four six-hour window fixed effects.

We use three measures of the gap G_{jm} between the searched and the booked itinerary. For each booking we compute the distance between the searched and the booked origin, g^o , and between the searched and the booked destination, g^d . Our main measure is their sum, $g^o + g^d$, averaged over the bookings of trip j and expressed in 100 km. An exact match yields $G_{jm} = 0$. For a near match, the measure captures the total displacement the traveler must cover outside the trip. To separate the role of the two endpoints, we also use the share of bookings of trip j with a gap at the origin and with a gap at the destination (a gap larger than 0.5 km). Finally, to allow for nonlinear effects, we split these shares into minor gaps (between 0.5 and 25 km) and severe gaps (more than 25 km).

Consumer heterogeneity enters through age and gender. We interact price with age categories (30–49, 50 and above, and not reported, with 10–29 as the base), so that price sensitivity can vary across generations. We also interact the bus and carpool dummies with age and gender, so that preferences for each mode can differ across demographic groups. These interactions are estimated in the first stage from the booking-level data.

In an additional specification, we control for the size of the market on both sides of the platform: the number of searches on the route and date (demand), the number of buses or carpool drivers offering the searched route (supply), and the number of results of each mode displayed to the searcher. These variables enter in logs, $\ln(1 + x)$, as exogenous controls.

Identification. We do not observe ξ_{jm} but consumers and drivers do. This leads to endogeneity issues as prices are likely to respond to unobserved shocks in ξ_{jm} . For instance, if prices are positively correlated with unobserved quality, price coefficient α will be downward biased (in absolute value).

To address this issue, we rely on instrumental variables targeting price identification. More specifically, we use the variable cost per seat as the main cost shifter. For trip j of transportation mode k , departing from origin o on date t , the cost is given by

$$c_{jt} = \frac{\frac{p_{o(j)t}^f d_j}{\eta_{k(j)}} + wa_{k(j)} T_j + \tau_{k(j)} d_j}{C_{k(j)}},$$

where $p_{o(j)t}^f$ is the daily fuel price at origin o (EUR/liter), d_j is the trip distance in km, and T_j is the travel time in hours. We assume that, regardless of the transportation mode, drivers always refuel before starting the journey, and buses use diesel and carpools use gasoline (E10). $\eta_{k(j)}$ is a mode-specific measure of fuel efficiency which is set to 3.33 and 14 km per liter for buses

and cars, respectively. The driver’s labor cost is represented by $wa_{k(j)}T_j$, which is 18 EUR/h for buses and zero for carpooling services (the driver makes the trip regardless of passengers). $\tau_{k(j)}$ is the cost per kilometer associated with road tolls. Finally, we divide total cost by total capacity $C_{k(j)}$, which is the total number of seats offered.¹⁰ Since the variation of the cost shifter comes from price dispersion across locations and daily fuel prices, we add the cost per seat per kilometer as a second cost shifter, and interact all instruments with mode dummy.¹¹ As robustness, we add a Hausman-type instrument. We use the average price of the same mode on the same date on routes that share no endpoint with market m .

The search algorithm of the platform provides consumers with a list of trips, which may include both exact-match and near-match trips. The latter may exhibit different unserved parts of the requested route, leading to distinct gap distances. Therefore, the gap coefficient is identified from variation arising from the distance gap across booked trips.

Two features of the setting support this assumption. First, a near-match trip in market m is an exact-match trip in some other market $m' \neq m$, i.e., the driver posted the near-match trip to serve route m' . Its existence, departure time and price respond to demand conditions on m' , captured by $\xi_{jm'}$, and the trip reaches consumers in m only because the algorithm displays it there. The assumption would fail if demand shocks on m and m' were correlated on the same date after the fixed effects absorb route-specific and calendar components. Second, drivers cannot target near-match markets. The platform does not disclose how it forms and ranks the list of results, which depends on the distance to the searched endpoints and on other trip attributes such as driver rating, driver experience and vehicle type. The attributes a driver chooses are therefore not set in anticipation of appearing in market m .

A remaining concern is that G_{jm} is computed from realized bookings. For instance, a negative demand shock might reduce the offer of exact-match trips, increase average gap and lead to fewer bookings. We address this by controlling for the number of offers on the searched route and the number of results displayed to the searcher.¹²

5.3 Estimation

In the estimation of the demand model, we make a series of assumptions for computational simplicity. First, within a market, we aggregate trips to the mode (i.e. inter-city bus or carpooling) and departure window level. The windows are defined based on six-hour time

¹⁰Similar costs have been used in other work to measure, e.g., distribution costs (De Loecker and Scott, 2026). The bus labor cost accounts for the French road-transport collective agreement and plus social contributions.

¹¹In Appendix B we report the distribution of fuel prices between 2021 and 2022 across France.

¹²This can still be problematic if the number of offers and results displayed respond to local-specific shocks (e.g., regional celebrations) that are not captured by route and month fixed effects.

intervals, dividing the day in 4 different windows. As a result, a typical consumer in market m chooses between 9 alternatives, including 8 inside goods plus the outside option. We aggregate the distinct variables accordingly (i.e. prices, distances, etc.). Second, we abstract from the dynamics related to purchasing horizon and the implied choice set restriction across consumers. We assume that consumers make decisions at the same time and observe all aggregate trips in the market.

Third, a market contains up to eight different trips (two modes times four departure windows) so the variance-covariance matrix adds up to 36 parameters to the estimation. The high-dimensionality of the parameter space and the nonlinearity nature of the problem pose computation challenges. For this reason, we restrict the covariance of the unobserved taste shocks to a factor structure. Specifically, the error of trip j for consumer i is

$$\varepsilon_{ijm} = \sigma_P \nu_i^P + \sigma_C \nu_{ic(j)}^C + \sigma_W \nu_{iw(j)}^W + \eta_{ijm},$$

where $w(j)$ is the departure window of trip (j), $c(j)$ denotes the mode, and all ν and η are independent standard normal. ν^P is a platform shock common to all inside alternatives, ν^C is a carpool-specific shock, and ν^W is a shock to specific departure window. Under this restriction, any two trips are correlated through the online platform σ_P^2 , plus σ_C^2 if both are carpooling services, plus σ_W^2 if they share the departure window. The resulting variance-covariance matrix of inside alternatives depends on only three parameters ($\sigma_P, \sigma_C, \sigma_W$).¹³ Finally, we conduct the analysis using a sample of markets. Specifically, we select the routes with the most bookings and for each of each route, we select randomly up to 24 departure dates spread across months and weekdays. The final sample contains 1,458 markets that correspond to 89 routes over 13 months. There are in total 9,958 products, 169 thousand bookings and about 1.68 million searches. The latter is used as a measure of market size.

Following Goolsbee and Petrin (2004), we employ an estimation algorithm that consists of two stages. First, we maximize the likelihood function using consumer-level data and including the separate trip-market fixed effects. This identifies all the parameters governing consumer heterogeneous valuations that are included in μ_{ijt} . In the second stage, we project the estimated fixed effects δ_{jm} onto price and trip characteristics, accounting for the respective endogeneity of prices.

In the first stage, given a value of the parameters $\theta = (\alpha, \beta, \sigma)$ and the vector of fixed effects δ , the probability that consumer i purchases trip j is given by (we omit the market index for

¹³The utility of the outside option is $u_{i0m} = \eta_{i0m}$ and the unit variance of η fixes the scale. In the current analysis, only carpool trips receive a mode-specific shock.

simplicity)

$$\begin{aligned} s_j(\theta, \sigma; z_i) &= s_{ij} \\ &= \int_{A_{ij}} dF(\varepsilon_i), \end{aligned}$$

Where A_{ij} is the set of ε_i such that trip j provides the highest utility. We numerically integrate out the multivariate normal errors using a frequency simulator with 4000 draws per consumer. The corresponding likelihood function is then given by

$$L = \prod_{i=1}^I \prod_{j=1}^J s_{ij}^{j(i)},$$

where j is an indicator function taking the value of one if consumer i chooses trip j , and zero otherwise. In the spirit of [Berry et al. \(1995\)](#) and following [Goolsbee and Petrin \(2004\)](#), we restrict the search over the space of θ , and solve for the vector $\delta(\theta)$ that matches observed and predicted market shares.¹⁴ We search for the parameter values θ that maximize the likelihood function. If the parameters do not maximize the objective function, we make a new guess and repeat the search routine.

In the second stage, we use the estimate fixed effects $\delta(\theta)$ and project them onto price and trip characteristics using a two-stage least square estimator. To address the correlation between ξ_{jm} and price p_{jm} , we use the cost per seat and its interaction with fuel price as cost shifters.

6 Estimation Results

This section examines how consumers choose among transportation modes on BlaBlaCar France and quantifies the extent to which the presence of near-match trips, and the implied gap distance – distorts choices. Following the structure of the empirical model, we present the results in two stages. Additional robustness checks are reported in Appendix C.2.

First-Stage Estimates. Table 7 reports the estimates of consumer heterogeneity from the Simulated Maximum Likelihood Estimation. Specification (1) includes interactions between consumer’s age categories and price, using group 10-29 as the base category. Specification (2) adds an interaction between bus dummy and gender. Specification (3) adds full interactions between demographics (age and gender) and transportation modes. Below, we focus the analysis on specification (3) as a likelihood-ratio test reject the restricted specifications.

¹⁴We use a nonlinear least squares routine to solve, market by market, for the $\delta(\theta)$ that minimizes the squared difference between predicted and observed market shares. See footnote 27 in [Goolsbee and Petrin \(2004\)](#) for further details.

The results show the valuation of trip characteristics relative to the base group (male aged 10-29). The price coefficients imply that consumers aged 30-49 are less price sensitive than younger consumers. The point estimate for consumers aged 50 and above is similar but imprecisely estimated. Females obtain greater utility from carpooling trips, while their valuation for buses is similar to the male group. Interestingly, the valuation of both transportation modes shows an inversed-u relationship with respect to age categories. The valuation peaks for the middle age group (age 30-49) and decreases at extremes ages. Moreover, consumers aged 30 and above value carpooling more than bus trips.

When consumers search for trip, they might have privacy concerns and opt not to reveal their demographics. These privacy-sensitive consumers exhibits the lowest price sensitivity among all consumers. This seems to suggest that consumers are willing to pay a higher price so long as their identity (e.g. demographics) remain private. Additionally, these types of consumers get higher utility from buses and lower utility from carpooling. This finding has two possible explanations. On the one hand, privacy concerns are more prevalent in peer-to-peer services, where passengers share a ride with fewer people or even with only the driver. Thus, privacy-sensitive consumers might be less willing to choose this service. On the other hand, long-distance carpooling drivers may be less likely to accept requests from passengers whose profile reveals little information. The not-reporting demographics is less important for inter-city buses which explains why privacy-sensitive passengers have a positive valuation for this mode.

Table 7
FIRST STAGE PARAMETER ESTIMATES

	(1)	(2)	(3)
<i>Price × demographics</i>			
Price × age 30–49	−0.001 (0.002)	0.000 (0.000)	0.007 (0.001)
Price × age 50+	0.002 (0.004)	−0.010 (0.005)	0.007 (0.006)
Price × age unreported	0.020 (0.001)	0.022 (0.001)	0.008 (0.001)
<i>Bus × demographics</i>			
Bus × female		−0.754 (0.169)	0.003 (0.002)
Bus × age 30–49			0.015 (0.003)
Bus × age 50+			−0.051 (0.006)
Bus × age unreported			0.250 (0.017)
<i>Carpool × demographics</i>			
Carpool × female			0.164 (0.031)
Carpool × age 30–49			0.089 (0.030)
Carpool × age 50+			0.024 (0.006)
Carpool × age unreported			−0.867 (0.106)
<i>Taste heterogeneity (std. dev.)</i>			
Platform shock	0.858 (0.014)	1.692 (0.008)	0.746 (0.011)
Carpool shock	0.181 (0.016)	0.051 (0.004)	0.633 (0.019)
Departure-window shock	0.712 (0.015)	0.534 (0.017)	0.531 (0.012)
Log-likelihood	−266,855.7	−264,303.5	−257,696.3
Observations	169,052	169,052	169,052

Notes: Simulated maximum likelihood conditional on booking. A product is a cell (searched route × departure date × mode × six-hour departure window); a market is a route × date, and the cell-specific mean utilities are concentrated out by matching the observed cell shares; the price level and the other cell characteristics enter through them, so the price interactions are deviations of the price coefficient from the base group. Base demographic group: age 10–29, male. Taste heterogeneity: normal shocks common to all platform alternatives, to the carpool alternatives, and to the alternatives of a departure window (standard deviations reported; idiosyncratic variance normalised to one). Sample: 89 routes, 1,458 markets, 9,958 cells, 169,052 bookers (the observations) and 1,676,719 searches; 4,000 simulation draws. Standard errors (GMM sandwich) in parentheses.

The parameters of the restricted variance-covariance matrix are reported at the bottom of

Table 7. These point estimates are precisely estimated, even across specifications. These results indicate that unobserved heterogeneous tastes are correlated among high-level attributes such as BlaBlaCar platform, carpooling services and departure window. If a morning carpooling service is removed from the market, consumers will switch towards other morning services offered by the platform, before they consider other departure windows or rivals' services (outside option).

Second-Stage Estimates. Table 8 displays the estimates of the preferences that are common across consumers, obtained from the regression of δ_{jm} on trip characteristics. The first two columns report OLS estimates with mode-specific price coefficients (column 1), and with a single price coefficient (column 2). Columns (3)-(6) show two-stage least squares estimates with different specifications. The specifications in columns (3) and (4) are the 2SLS counterparts of the OLS columns, whereas columns (5) and (6) introduce alternative gap measures and control for additional characteristics. All specifications control for month, weekday, departure window and route (OD) fixed effects.

Table 8
STAGE TWO ESTIMATES FROM FIXED EFFECTS REGRESSIONS

	OLS		2SLS (IV11)			
	(1)	(2)	(3)	(4)	(5)	(6)
Price $\times$ bus	−0.008 (0.001)		−0.022 (0.004)			
Price $\times$ carpool	−0.008 (0.001)		−0.025 (0.004)			
Price		−0.008 (0.001)		−0.018 (0.004)	−0.023 (0.004)	−0.036 (0.005)
Gap sum (100 km)	−0.198 (0.007)	−0.198 (0.007)	−0.202 (0.008)	−0.199 (0.008)		
Origin gap, minor					−0.052 (0.040)	0.035 (0.047)
Origin gap, severe					−0.732 (0.035)	−0.544 (0.053)
Destination gap, minor					−0.004 (0.039)	0.064 (0.047)
Destination gap, severe					−0.749 (0.035)	−0.498 (0.058)
Log searches						−0.009 (0.060)
Log offers						0.211 (0.022)
Log results displayed						0.299 (0.021)
Departure 0–6h	−0.931 (0.019)	−0.931 (0.019)	−0.933 (0.019)	−0.932 (0.019)	−0.939 (0.019)	−0.996 (0.022)
Departure 12–18h	0.105 (0.016)	0.105 (0.016)	0.089 (0.017)	0.094 (0.017)	0.088 (0.017)	0.071 (0.018)
Departure 18–24h	−0.512 (0.015)	−0.512 (0.015)	−0.493 (0.017)	−0.496 (0.017)	−0.502 (0.017)	−0.539 (0.020)
Bus	−0.008 (0.025)	−0.023 (0.013)	−0.077 (0.030)	−0.003 (0.014)	−0.053 (0.018)	0.892 (0.043)
Constant	−3.182 (0.176)	−3.176 (0.175)	−2.964 (0.158)	−3.046 (0.162)	−2.775 (0.133)	−4.419 (0.361)
Observations	9,533	9,533	9,533	9,533	9,533	8,630
F-stat			70.4 / 72.8	29.1	29.2	18.9

Notes: Stacked regressions of the cell mean utilities $\hat{\delta}$ from stage 1 (column (3) of Table 7). Cell: searched route $\times$ date $\times$ mode $\times$ six-hour departure window; market: route $\times$ date. Sample: 89 routes, 1,458 markets; observations are the cells with complete data (fewer with network variables, which are missing for some markets). Price: transaction price in euros (carpool: driver amount plus commission and VAT); Price $\times$ bus and Price $\times$ carpool are mode-specific coefficients. Gaps: distance between the searched and the booked origin (destination). Gap sum: mean origin plus destination gap of the cell's bookings, in 100 km. Gap bands: shares of the cell's bookings with a minor (0.5–25 km) or a severe (>25 km) gap; base: no gap. Network variables, in $\ln(1+x)$ and treated as exogenous: searches on the route-date, offers of the mode for the searched origin–destination, results displayed for the mode. Departure dummies relative to 6–12h; Bus: bus dummy (carpool base). Fixed effects: month, weekday, route (departure window reported). Instruments, interacted with the mode dummy: IV11 = trip cost per seat and per seat-km (fuel, driver time and tolls over seats; bus on diesel, carpool on E10 prices). F-stat: heteroskedasticity-robust first-stage F on the excluded instruments; with two prices, Sanderson–Windmeijer conditional F for the bus / carpool price (not robust), because the unconditional F is mechanically inflated: each mode's price is zero on the other mode's cells, which the mode-interacted instruments fit exactly. Heteroskedasticity-robust standard errors in parentheses.

In Section 5.2, we argued that ignoring the correlation between prices and unobserved trip quality biases the price sensitivity parameter towards zero. In line with this, the OLS estimates in columns (2) and (3) show a negative and significant but small price coefficient. The remaining columns address the endogeneity problem by using instruments based on the cost-per seat variable. The price coefficient increases in magnitude from -0.008 to -0.018 in column (3), and to around -0.024 in column (4).¹⁵ Column (3) shows that the mode-specific price coefficient are slightly different, with a larger price sensitivity for carpooling. Although a Wald test rejects their equality, the difference is rather small, and for parsimony we continue the analysis with a single price coefficient. These price coefficients determine the substitution patterns among the inside goods, and between inside goods and the outside option. In Appendix C.1, Table 18 and Figure 15 show the implied own- and cross-price elasticities, and the corresponding diversion ratios.

The estimates for departure windows show that consumers prefer to travel during the day, especially in the early afternoon (12-18h), relative to the morning (6-12h). The least valuable time for travelling is before 6h, followed by the evening journeys (18-24h). The bus estimate is negative. Conditional on prices and gaps, this suggest that on average consumers value carpooling slightly more than inter-city buses. The sign of this coefficient, however, reverses once we control for the number of offers and the number of results displayed.

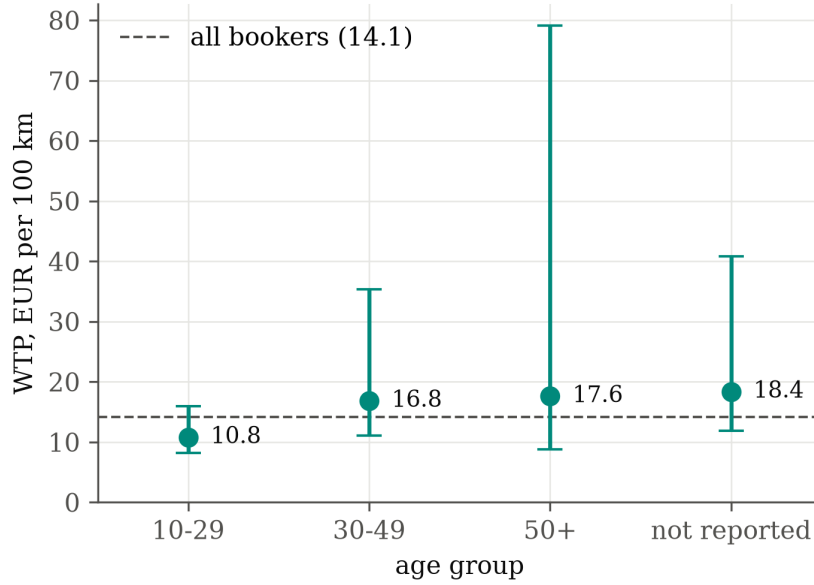
The gap coefficient is negative and precisely estimated. This shows that the mismatch between the searched and booked itinerary generates disutility for consumers, which highlights the potential benefits of multimodal technology. Column (5) splits the gap variable into four components: minor and sever gaps at both the origin and destination. We find that consumers are indifferent about minor mismatches at any endpoint, but obtain a strong disutility from severe gaps. These results hold after controlling for the number of offers and the number of displayed searches (column (6)). In this last specification, the severe gaps associated with departure location yields higher disutility than the ones of destination location.

These demand estimates imply that consumers dislike the mismatch between searched and booked trips, and that they are willing to pay to get rid of the uncovered part of the journey. To study their willingness to pay (WTP) for gap reduction, use the estimated gap coefficient of specification (4) in Table 8 and the heterogeneous price sensitivity parameter obtained from both estimation stages. Figure 7 shows the results across age categories.

On average, consumers would be willing to pay 14.1 EUR to reduce the gap by 100 kms.

¹⁵Table 21 in the Appendix presents the results of the first-stage regression of price on the instrumental variables. The bottom of Table 8 reports the resulting F-statistic for the excluded instruments. The F-statistic exceeds 18 in all specifications and we can reject the null that the instruments are not relevant.

Figure 7
Willingness-to-Pay for Origin-Destination Gap Reduction



Notes: Willingness to pay, in euros per 100 km of gap (origin plus destination gap). The graph shows the point estimates and the 5–95% range of parametric bootstrap (10,000 draws).

The heterogeneity in price sensitivity results in different WTP across age categories. While consumers aged 10-29 would pay less than 11 EUR for the reduction of the gap, older consumers would pay between 56 and 63 percent more for the same gap reduction. The privacy-sensitive consumers (i.e. unreported) show the highest WTP 18.4 EUR.¹⁶ Mapping the estimates to the data, consumers are willing to pay 6.95 EUR for the average distance gap of 49.3 km, roughly 34% of the average price for inter-city buses and 29% of the average price for carpooling. These results suggest that there are substantial gains from multimodal technology, and its implementation – with associated fees – should not steer consumers away from the market. We investigate this further in the next section.

7 Understanding the Market Frictions

7.1 Welfare Analysis

The demand estimates show that consumers dislike the gap between the trip they search for and the trip they book, and are willing to pay to reduce it. We use the estimated demand model to assess the potential welfare gains from multimodal technology. We consider counterfactual scenarios in which the technology reduces the distance between searched and booked itineraries, and examine the resulting changes in consumer surplus and booking probabilities. We then

¹⁶The wide confidence bands for the age 50 and above reflects the imprecise price coefficient for this group of consumers.

introduce a uniform surcharge per booked trip to assess how much the platform could charge before eliminating the average consumer-surplus gains from multimodal integration.

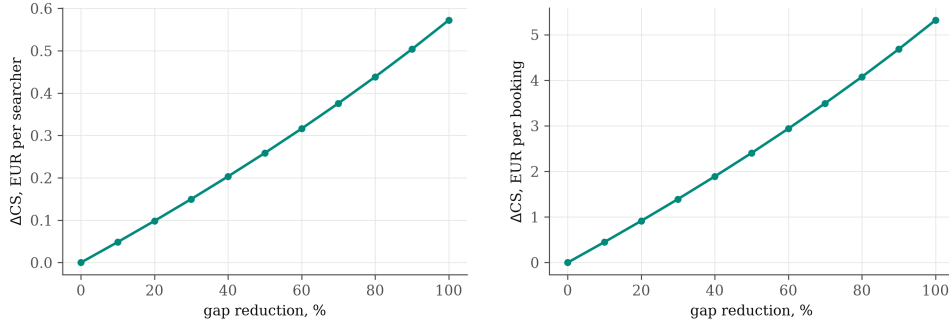
Two features of this exercise should be kept in mind. First, these are demand-side, partial-equilibrium counterfactuals: prices and the set of offered trips remain at their observed values, and neither the platform nor transport providers adjust their decisions in response to the technology. The results therefore capture the consumer-welfare channel. We do not quantify the responses in producer surplus or the environmental effects that may arise if multimodal integration changes the use of private cars, congestion, or emissions. Second, consumer surplus is measured as compensating variation, calculated from the inclusive value of the estimated demand model and converted into euros using the estimated price coefficient. We report the results both per searcher and per booking.

Consumer surplus from gap reduction. Figure 8 shows the change in consumer surplus for different degrees of gap reduction, measured both per searcher and per booking. The gains increase as the gap is reduced. Eliminating the gap entirely increases consumer surplus by approximately EUR 0.57 per searcher and EUR 5.6 per booking. A 50% reduction yields gains of approximately EUR 0.26 per searcher and EUR 2.4 per booking. The difference between these measures reflects the relatively low booking rate. The per-searcher measure captures the average welfare gain across all searches, while the per-booking measure focuses on searches that result in a booking. Since only around 11% of searchers book, the estimated gain per booking is substantially larger than the gain per searcher. Applying the estimated gain per searcher to the 18.3 million searches in the sample implies an increase in consumer surplus of approximately EUR 10.4 million under full gap elimination. To put these magnitudes in perspective, the gain of EUR 5.6 per booking is about one quarter of the average fare on the platform (roughly EUR 20–24), and the EUR 10.4 million aggregate accrues from a change that adds no new physical supply.

These results are consistent with the demand estimates, which indicate that consumers experience substantial disutility from large gaps between searched and booked itineraries, while smaller gaps have little effect on utility. Reducing the gap therefore makes available trips more attractive, with larger reductions removing more of the spatial mismatch that consumers find costly. Overall, the counterfactual suggests that reducing these gaps can generate meaningful consumer-surplus gains.

Market expansion. The gains above reflect not only increases in the surplus of consumers who book in the observed market, but also the entry of additional bookings. Figure 9 shows

Figure 8
Change in Consumer Surplus from Reducing the Gap



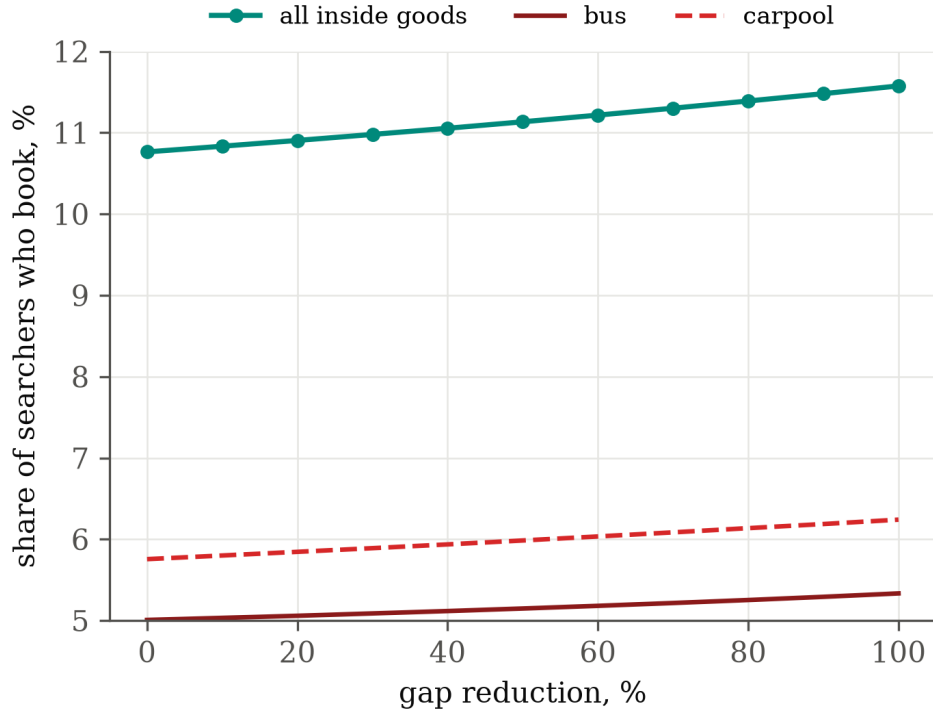
Notes: The gap of every product is reduced by the share on the horizontal axis, holding prices fixed (0% = observed data). Left: change in consumer surplus per searcher. Right: total change divided by the observed number of bookings.

how gap reduction affects booking probabilities. The overall booking rate increases from approximately 10.75% in the observed scenario to 11.6% when the gap is fully eliminated. This represents an increase of around 0.85 percentage points, or approximately 8% relative to the initial booking rate. Thus, the potential benefits of multimodal technology extend beyond consumers who already book: reducing the gap also induces some consumers who would otherwise choose the outside option to book a trip. The increase is larger for carpooling, whose booking share rises from approximately 5.75% to 6.25%, than for buses, whose share increases from 5.0% to 5.33%. Carpooling therefore accounts for the larger share of the additional bookings generated by the counterfactual.

This pattern is consistent with the reduced-form evidence in Section 4.3, which suggests a stronger relationship between endpoint connectivity and the propensity to accept a near-match for carpooling. It is also consistent with the broad geographic coverage of the carpooling network: reducing the gap can make more of its existing trips attractive to passengers whose searched itineraries are not directly served. However, the counterfactual does not identify which transport mode these additional passengers would otherwise have used. The increase in bookings therefore indicates the potential for multimodal integration to expand the use of services on the platform, but does not by itself establish that private-car use would decline or that congestion and emissions would fall.

Taken together, these results show that reducing spatial gaps can increase consumer surplus and expand the number of bookings. The welfare gains arise both because trips become more attractive to consumers who already book and because some consumers who previously chose the outside option are induced to book. The extent to which this market expansion translates into broader social benefits depends on how these additional bookings affect travel choices outside the platform.

Figure 9
Extensive Margin and Gap Reduction



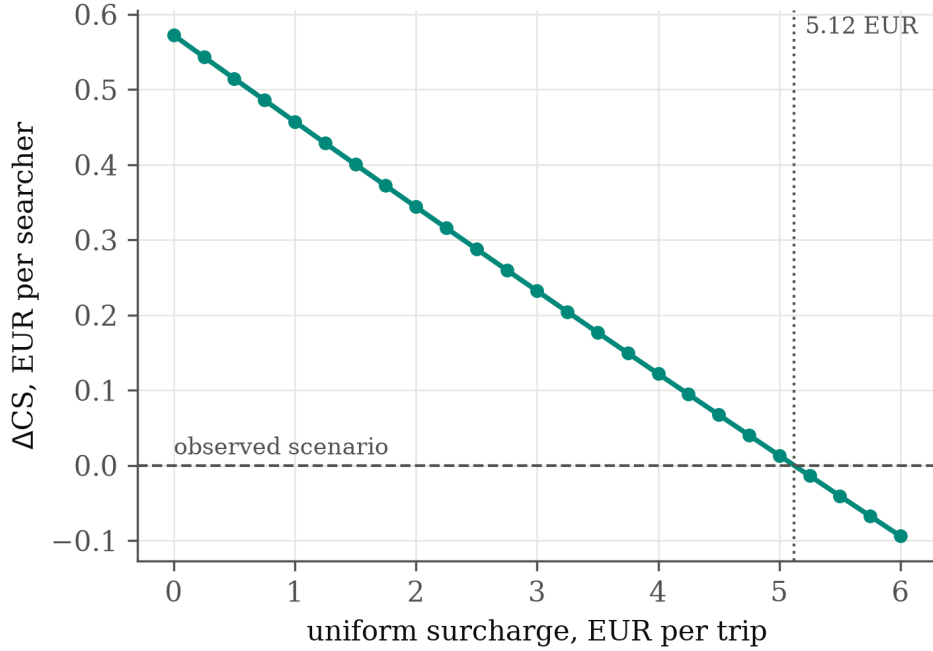
Notes: The figure plots the share of searchers who book (one minus the outside-option share), in percent, as the origin-plus-destination gap of every product is reduced by the share on the horizontal axis (0% = observed market), holding prices fixed. The solid line with markers is all inside goods (bus plus carpooling); the remaining lines are bus and carpooling separately. Counterfactual based on the estimated model.

Pricing the technology. The multimodal technology may entail costs, raising the question of whether a fee could be charged without eliminating the consumer-surplus gains it generates. We therefore assess how a uniform surcharge per booked trip affects consumer surplus when the gap is fully eliminated.

Figure 10 reports the change in consumer surplus relative to the observed scenario as the surcharge increases. Without a surcharge, eliminating the gap generates a gain of approximately EUR 0.57 per searcher. This gain declines as the surcharge increases and reaches zero at approximately EUR 5.12 per booked trip. Beyond this threshold, consumer surplus falls below its level in the observed scenario, suggesting that the surcharge more than offsets the average gains from reducing the gap.

The estimated break-even surcharge provides a benchmark for assessing the scope for charging for multimodal integration without reducing average consumer surplus relative to the observed scenario. In relative terms this threshold is large: EUR 5.12 is close to a quarter of the average fare and comparable to the platform's current per-trip commission, so the technology could be priced at a level that materially raises platform revenue while still leaving the average consumer

Figure 10
Uniform Surcharge with Full Gap Removal



Notes: The gap is fully removed and every booked trip is charged a uniform surcharge (horizontal axis). The dashed line marks the observed scenario (observed gaps and prices); the dotted line marks the surcharge at which consumer surplus returns to its observed level.

better off than in the observed market.

It should not, however, be interpreted as a fee that leaves every consumer better off. A uniform surcharge affects consumers differently depending on their preferences and price sensitivity, and may discourage some marginal consumers from booking. Moreover, the break-even surcharge is not directly comparable to the EUR 6.95 willingness to pay estimated for the average gap of 49.3 km. The latter measures willingness to pay for reducing a given gap, whereas the counterfactual measures the average welfare effect across all searches, including those for which the observed trip has no gap, and accounts for the effect of the surcharge on booking decisions. As documented in Section 6, a fee of around EUR 5 can leave the average consumer better off while reducing the net benefit for the youngest and most price-sensitive searchers, whose willingness to pay is below EUR 11. Other consumer groups, including the privacy-sensitive group reported in that section, have a higher willingness to pay, reaching EUR 18.4, and would retain a larger surplus.

Overall, the welfare analysis suggests that reducing spatial gaps can increase consumer surplus and booking probabilities, with the largest increase in bookings occurring for carpooling. It also suggests that a uniform surcharge can be introduced without eliminating the average consumer-surplus gain, provided it remains below the estimated break-even threshold. These estimates capture the demand-side benefits of multimodal integration. Assessing its overall market impact

would additionally require accounting for firms' pricing and supply responses, the resulting changes in producer surplus, and any environmental effects.

7.2 Discussion and Future Research

The preceding analysis shows that multimodal technology can generate consumer-surplus gains by reducing spatial gaps between searched and booked trips and increasing the share of searchers who book. However, the current framework does not account for the strategic responses of drivers, bus operators, or the platform, either in terms of entry and service provision or pricing. The estimated welfare gains should therefore be interpreted as demand-side effects under observed supply and pricing conditions, rather than as the equilibrium effects of introducing the technology.

These limitations are also relevant for policy. By connecting services across modes and expanding the set of itineraries available between origin-destination pairs, multimodal integration can improve access to shared and public transport. This may contribute to environmental policy objectives by making alternatives to private-car travel more attractive and potentially reducing congestion and emissions. However, these effects depend on how consumers change their travel choices. If additional bookings on the platform replace private-car journeys, multimodal integration may generate environmental benefits. If they instead replace existing bus or carpooling trips, or induce additional travel, the environmental effects may be more limited. Quantifying these effects requires accounting for the transport modes consumers would otherwise have chosen, as well as the responses of transport providers. The current analysis establishes the potential consumer benefits of reducing spatial gaps, but does not directly measure changes in private-car use or emissions.

A first direction for future research is to account for supply responses and the network effects that characterize two-sided transportation markets. Combining multiple transport modes can create additional matching opportunities, extend network coverage, and improve service quality. These changes may increase passenger demand and, in turn, strengthen drivers' incentives to enter the platform and bus operators' incentives to expand their services. However, an increase in supply does not necessarily translate into higher matching probabilities. For example, if driver entry grows disproportionately relative to demand, competition for passengers may intensify and the probability of matching may decline. More generally, the interaction between passenger demand, supplier entry, and matching outcomes can either reinforce or offset the initial benefits of multimodal integration. Pricing decisions introduce an additional channel, since the platform may adjust its fees in response to changes in demand and supply, affecting both participation

and consumer surplus. Understanding these effects requires a framework that jointly accounts for demand, supplier entry, and the matching process.

Our future research will address these questions by developing and estimating a structural matching model using proprietary BlaBlaCar data on ride requests, bookings, and BlaBlaBus usage. In this framework, drivers and passengers decide whether to participate in the platform based on the expected value of matching with the other side of the market. The model will allow us to study how multimodal integration affects entry decisions, matching probabilities, and booking outcomes, accounting for feedback between demand and supply. It will also provide a framework for evaluating how changes in service quality and platform pricing affect market outcomes and consumer welfare. By incorporating these supply-side responses, this approach can assess whether the gains identified in the current analysis persist once drivers and the platform adjust their decisions, and whether multimodal integration improves the overall efficiency of the market.

A second limitation concerns the timing of consumer searches and the availability of alternatives. The choice set faced by a passenger depends not only on the origin and destination of the trip but also on when the search takes place. For a given departure date and origin-destination pair, the available alternatives may differ depending on how far in advance the passenger searches. Although our estimation partially accounts for variation in available supply, it does not fully model how choice sets evolve over time. This may affect the estimated preferences and counterfactual results, since multimodal integration can change the alternatives available to passengers at different points in time. The direction of the resulting bias is not determined a priori: depending on how availability evolves, ignoring these dynamics may either overstate or understate the welfare gains.

Future research will therefore also account for inter-temporal variation in choice sets and the possibility that consumers adjust the timing of their booking decisions. In transportation markets, consumers may choose among the alternatives available at the time of the search or wait for potentially better options to emerge. Waiting, however, entails a trade-off: new services, such as last-minute carpooling trips, may become available, while existing alternatives may sell out. We plan to develop a structural model of dynamic consumer choice in which individuals face heterogeneous choice sets and make forward-looking booking decisions under uncertainty about the availability and quality of future alternatives. This framework will allow us to distinguish the effects of multimodal integration on the set of available trips from its effects on consumers' booking decisions over time. It will also provide a more complete assessment of how the technology affects consumer welfare when both the composition and the timing of available

alternatives can change.

8 Conclusion

In this paper, we quantify consumers’ valuation of multimodal technologies using detailed data from a long-distance transportation platform in France. We first document and quantify the market frictions associated with spatial mismatches between searched and booked trips. We then develop and estimate an empirical model of consumers’ transportation-mode choices to assess the potential welfare gains from the introduction of a multimodal technology that allows passengers to combine transport modes within a single itinerary.

Our results show that reducing spatial mismatches can generate substantial consumer-surplus gains. Consumers are willing to pay for a technology that reduces the gaps between their searched and booked itineraries, with an estimated break-even surcharge of approximately EUR 5.12 per booked trip. The benefits extend beyond passengers who already make a booking. By allowing passengers to combine transport modes and avoid covering the unserved portions of their journeys themselves, multimodal integration can also make the platform attractive to consumers who would otherwise leave without booking. Consistent with this mechanism, our counterfactual analysis shows that the share of searchers who book increases from 10.75% to 11.6%, with a larger increase in carpooling than in bus bookings. Finally, the welfare gains are mainly driven by the reduction of severe spatial mismatches, while consumers are largely indifferent to minor gaps. This highlights the importance of multimodal integration in making trips feasible when the available services do not cover substantial portions of the desired itinerary.

Taken together, these findings suggest that multimodal integration can improve consumer welfare through two complementary channels: increasing the value of trips that can already be booked and expanding the set of consumers who find a suitable trip. The results also indicate that the value of the technology depends on the extent of the spatial mismatches it alleviates. However, our counterfactuals hold prices and service availability fixed and therefore capture demand-side effects under the observed market conditions, rather than the full equilibrium effects of introducing the technology.

Future research should address two main limitations. First, the current framework focuses on passenger demand and does not account for the strategic responses of drivers, bus operators, or the platform. Introducing a structural model of the matching process between drivers and passengers would make it possible to study how improved service quality affects participation and entry on both sides of the market, and how the resulting supply responses influence matching probabilities and consumer welfare. This is particularly important because multimodal inte-

gration may change the relationship between buses and carpooling: rather than acting only as substitutes, the two modes may also become complements when they can be combined within a single itinerary. Accounting for these interactions would help establish whether the demand-side gains identified in this paper persist once providers and the platform adjust their decisions.

Second, further work could incorporate the evolution of available trips over time and consumers' decisions about when to book. The alternatives available to a passenger depend on when a search takes place, and consumers may face a trade-off between booking an available trip and waiting for potentially better alternatives. A dynamic model could capture these decisions and provide a more complete assessment of the welfare effects of multimodal integration.

More broadly, combining demand estimation with a model of supply, matching, and booking decisions would allow future research to evaluate how multimodal integration affects market outcomes under alternative pricing strategies. In particular, it would make it possible to study how changes in network coverage and consumer costs affect participation, market shares, and the platform's pricing incentives. This would provide a more comprehensive assessment of the potential effects of multimodal technologies beyond the demand-side gains quantified in the current paper.

Bibliography

- AGENCE DE L'ENVIRONNEMENT ET DE LA MAÎTRISE DE L'ÉNERGIE (ADEME) (2022): “BLABLAMODES : MaaS moyenne longue distance covoiturage-bus-train,” Retour d'expérience 5319, ADEME, accessed: 2026-10-13.
- ASTIER, N., P.-F. BOUQUET, AND X. LAMBIN (2023): “Riding together: eliciting travelers' preferences for long-distance carpooling,” *Available at SSRN 4360029*.
- BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): “Automobile prices in market equilibrium,” *Econometrica: Journal of the Econometric Society*, 841–890.
- BIAN, B. (2018): “Search frictions, network effects and spatial competition: Taxis versus Uber,” .
- BUCHHOLZ, N. (2022): “Spatial equilibrium, search frictions, and dynamic efficiency in the taxi industry,” *The Review of Economic Studies*, 89, 556–591.
- BUCHHOLZ, N., L. DOVAL, J. KASTL, F. MATĚJKA, AND T. SALZ (2020): “The value of time: Evidence from auctioned cab rides,” .
- CASTILLO, J. C. (2025): “Who benefits from surge pricing?” *Econometrica*, 93, 1811–1854.
- CASTILLO, J. C., D. KNOEPFLE, AND E. G. WEYL (2025): “Matching and pricing in ride hailing: Wild goose chases and how to solve them,” *Management Science*, 71, 4377–4395.
- CHEN, M. K., P. E. ROSSI, J. A. CHEVALIER, AND E. OEHLSEN (2019): “The value of flexible work: Evidence from Uber drivers,” *Journal of political economy*, 127, 2735–2794.
- CHRISTENSEN, P., G. NINO, AND A. OSMAN (2023): “The demand for mobility: Evidence from an experiment with uber riders,” Tech. rep., National Bureau of Economic Research.
- COHEN, P., R. HAHN, J. HALL, S. LEVITT, AND R. METCALFE (2016): “Using big data to estimate consumer surplus: The case of uber,” Tech. rep., National Bureau of Economic Research.
- DE LOECKER, J. AND P. T. SCOTT (2026): “Markup estimation using production and demand data: An application to the us brewing industry,” *Review of Economic Studies*, rdag109.
- DELHOMME, P. AND A. GHEORGHIU (2016): “Comparing French carpoolers and non-carpoolers: Which factors contribute the most to carpooling?” *Transportation Research Part D: Transport and Environment*, 42, 1–15.

- EUROPEAN COMMISSION (2017): “Commission Delegated Regulation (EU) 2017/1926 supplementing Directive 2010/40/EU with regard to the provision of EU-wide multimodal travel information services,” Updated by Delegated Regulation (EU) 2024/490.
- (2023): “Proposal for a Regulation on Multimodal Digital Mobility Services (MDMS),” European Commission Legislative Initiative, framework establishing FRAND terms for cross-border transit booking apps.
- EUROPEAN PARLIAMENT (2026): “Common transport policy: overview,” Fact Sheets on the European Union, accessed: 2026-10-13.
- EUROPEAN PARLIAMENT AND COUNCIL OF THE EUROPEAN UNION (2010): “Directive 2010/40/EU on the framework for the deployment of Intelligent Transport Systems in the field of road transport and for interfaces with other modes of transport,” Amended by Directive (EU) 2023/2661.
- FARAJALLAH, M., R. G. HAMMOND, AND T. PÉNARD (2019): “What drives pricing behavior in peer-to-peer markets? Evidence from the carsharing platform BlaBlaCar,” *Information Economics and Policy*, 48, 15–31.
- FRECHETTE, G. R., A. LIZZERI, AND T. SALZ (2019): “Frictions in a competitive, regulated market: Evidence from taxis,” *American Economic Review*, 109, 2954–2992.
- GAINEDDENOVA, R. (2022): “Pricing and efficiency in a decentralized ride-hailing platform,” *Unpublished working paper*.
- GHEORGHIOU, A. AND P. DELHOMME (2018): “For which types of trips do French drivers carpool? Motivations underlying carpooling for different types of trips,” *Transportation Research Part A: Policy and Practice*, 113, 460–475.
- GOEREE, M. S. (2008): “Limited information and advertising in the US personal computer industry,” *Econometrica*, 76, 1017–1074.
- GOOLSBEE, A. AND A. PETRIN (2004): “The consumer gains from direct broadcast satellites and the competition with cable TV,” *Econometrica*, 72, 351–381.
- HALL, J. D., C. PALSSON, AND J. PRICE (2018): “Is Uber a substitute or complement for public transit?” *Journal of urban economics*, 108, 36–50.
- HIDALGO, J. AND M. SOVINSKY (2023): “Internet (power) to the people: How to bridge the digital divide,” Tech. rep., University of Bonn and University of Mannheim, Germany.

- JACOBSON, S. H. AND D. M. KING (2009): “Measuring the potential for automobile fuel savings in the US: the impact of obesity,” *Transportation research part D: transport and environment*, 14, 6–13.
- LAGOS, R. (2000): “An alternative approach to search frictions,” *Journal of Political Economy*, 108, 851–873.
- MONCHAMBERT, G. (2020): “Why do (or don’t) people carpool for long distance trips? A discrete choice experiment in France,” *Transportation Research Part A: Policy and Practice*, 132, 911–931.
- SHAPIRO, M. H. (2018): “Density of Demand and the Benefit of Uber,” *mshapiro-jmp. pdf*.
- TARDUNO, M. (2021): “The congestion costs of Uber and Lyft,” *Journal of Urban Economics*, 122, 103318.
- YEUNG, T. Y.-C. AND D. ZHU (2022): “Intercity ridesharing to the rescue: Capacity flexibility and price stability of BlaBlaCar during the 2018 French railway strike,” *Transportation Research Part A: Policy and Practice*, 164, 270–290.
- YIN, B., L. LIU, N. COULOMBEL, AND V. VIGUIÉ (2018): “Appraising the environmental benefits of ride-sharing: The Paris region case study,” *Journal of cleaner production*, 177, 888–898.

A Institutional setting - the matching process

The matching process works as follows. First, at any time, a driver with empty seats on a given trip can make a post on the platform. She enters her departure time and location, as well as her arrival location. She may also fill in locations where she is willing to make a stop to drop off or pick up passengers on her way. She sets a price for the full trip and any section thereof if applicable. On the other side of the platform, prospective passengers can search rides by entering their desired date, departure and arrival cities. The platform in turn lists all the currently available rides matching these criteria. It highlights two results (the cheapest and the closest to the requested origin and destination), and then displays the full list of results in chronological order in terms of departure time. Passengers can finally book one or several seats on their preferred trip. The booking confirmation is either instantaneous or may pend approval by the driver if she opted in manual validation. Finally, the platform charges a commission fee of around 13/15% of the price requested by a driver, which is included in the price seen and paid by passengers.

On the main corridors between major cities, the platform also operates coach services. For these routes, search requests also return the available bus rides. A prospective passenger searching for a ride on the platform then faces a choice set composed of several carpooling rides and one or several bus rides.

B Willingness to multimode: full estimation tables

B.1 Conversion: $\Pr(\text{book})$ over all searches

Table 9

Conversion: Pr(book) over all searches — all bookings.

	(1) Pooled	(2) Pooled _×	(3) Split	(4) Split _×
<i>LPM (OD FE)</i>				
Connectivity (pooled)	+0.0198 (0.0175)	+0.0200 (0.0174)		
bus component			-0.0321 (0.0207)	-0.0317 (0.0215)
car component			+0.0413** (0.0183)	+0.0411** (0.0190)
Underserved		+0.0539*** (0.0190)		+0.0444 (0.0370)
Underserved × conn.		-0.0074** (0.0031)		
Underserved × conn. bus				-0.0027 (0.0097)
Underserved × conn. car				-0.0045 (0.0116)
Log direct supply	+0.0128*** (0.0009)	+0.0128*** (0.0009)	+0.0127*** (0.0008)	+0.0127*** (0.0008)
Price	-0.0003** (0.0001)	-0.0003** (0.0001)	-0.0002** (0.0001)	-0.0002** (0.0001)
<i>N</i>	20,559,924	20,559,924	20,559,924	20,559,924
<i>Logit</i>				
Connectivity (pooled)	-0.1821*** (0.0322)	-0.1807*** (0.0325)		
bus component			-0.1383* (0.0793)	-0.1361* (0.0815)
car component			-0.0487 (0.0845)	-0.0499 (0.0863)
Underserved		+0.2923 (0.3712)		+0.0957 (0.5521)
Underserved × conn.		-0.0402 (0.0618)		
Underserved × conn. bus				-0.0440 (0.1343)
Underserved × conn. car				+0.0152 (0.1518)
Log direct supply	+0.2131*** (0.0186)	+0.2131*** (0.0187)	+0.2086*** (0.0195)	+0.2086*** (0.0196)
Price	-0.0150*** (0.0028)	-0.0151*** (0.0028)	-0.0146*** (0.0028)	-0.0146*** (0.0028)
<i>N</i>	3,000,000	3,000,000	3,000,000	3,000,000
<i>Probit</i>				
Connectivity (pooled)	-0.0961*** (0.0165)	-0.0953*** (0.0166)		
bus component			-0.0752* (0.0422)	-0.0741* (0.0434)
car component			-0.0229 (0.0455)	-0.0234 (0.0465)
Underserved		+0.1605 (0.1956)		+0.0570 (0.2939)
Underserved × conn.		-0.0226 (0.0321)		
Underserved × conn. bus				-0.0217 (0.0718)
Underserved × conn. car				+0.0057 (0.0814)
Log direct supply	+0.1113*** (0.0093)	+0.1113*** (0.0093)	+0.1092*** (0.0095)	+0.1092*** (0.0096)
Price	-0.0073*** (0.0013)	-0.0073*** (0.0013)	-0.0071*** (0.0013)	-0.0071*** (0.0013)
<i>N</i>	3,000,000	3,000,000	3,000,000	3,000,000

Cluster-robust standard errors (origin) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Columns: (1) connectivity only; (2) adds the mismatch measure (underserved for conversion, gap for mismatch) and its interaction with pooled connectivity; (3) connectivity split into bus and car components; (4) split plus the mismatch measure and its split interactions. LPM (OD FE) absorbs origin, destination and day; LPM (region FE) uses region effects plus distance and market-size controls; logit and probit use month and day-of-week effects. Empty _× columns under probit reflect perfect separation of the gap measure.

Table 10

Conversion: Pr(book) over all searches — carpool bookings.

	(1) Pooled	(2) Pooled _×	(3) Split	(4) Split _×
<i>LPM (OD FE)</i>				
Connectivity (pooled)	+0.0007 (0.0094)	+0.0008 (0.0092)		
bus component			-0.0330** (0.0140)	-0.0314** (0.0138)
car component			+0.0241** (0.0120)	+0.0228** (0.0112)
Underserved		+0.0774*** (0.0216)		+0.0353 (0.0366)
Underserved × conn.		-0.0103*** (0.0033)		
Underserved × conn. bus				-0.0139** (0.0064)
Underserved × conn. car				+0.0048 (0.0093)
Log direct supply	+0.0106*** (0.0008)	+0.0106*** (0.0008)	+0.0105*** (0.0008)	+0.0105*** (0.0008)
Price	+0.0004*** (0.0001)	+0.0004*** (0.0001)	+0.0004*** (0.0001)	+0.0004*** (0.0001)
<i>N</i>	20,559,924	20,559,924	20,559,924	20,559,924
<i>Logit</i>				
Connectivity (pooled)	-0.4377*** (0.1134)	-0.4314*** (0.1156)		
bus component			-0.8538*** (0.1438)	-0.8571*** (0.1453)
car component			+0.3797** (0.1883)	+0.3878** (0.1909)
Underserved		+0.3535 (0.9253)		+0.6903 (0.7077)
Underserved × conn.		+0.0425 (0.1383)		
Underserved × conn. bus				+0.2866* (0.1489)
Underserved × conn. car				-0.2068 (0.1848)
Log direct supply	+0.3137*** (0.0307)	+0.3166*** (0.0302)	+0.2800*** (0.0295)	+0.2827*** (0.0293)
Price	-0.0041 (0.0039)	-0.0041 (0.0038)	-0.0000 (0.0043)	-0.0001 (0.0041)
<i>N</i>	3,000,000	3,000,000	3,000,000	3,000,000
<i>Probit</i>				
Connectivity (pooled)	-0.2160*** (0.0546)	-0.2125*** (0.0555)		
bus component			-0.4113*** (0.0692)	-0.4123*** (0.0697)
car component			+0.1874** (0.0912)	+0.1910** (0.0922)
Underserved		+0.2581 (0.4427)		+0.3864 (0.3499)
Underserved × conn.		+0.0060 (0.0657)		
Underserved × conn. bus				+0.1200 (0.0741)
Underserved × conn. car				-0.0972 (0.0928)
Log direct supply	+0.1500*** (0.0133)	+0.1513*** (0.0131)	+0.1370*** (0.0125)	+0.1383*** (0.0123)
Price	-0.0017 (0.0017)	-0.0017 (0.0017)	-0.0002 (0.0019)	-0.0002 (0.0019)
<i>N</i>	3,000,000	3,000,000	3,000,000	3,000,000

Cluster-robust standard errors (origin) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Columns: (1) connectivity only; (2) adds the mismatch measure (underserved for conversion, gap for mismatch) and its interaction with pooled connectivity; (3) connectivity split into bus and car components; (4) split plus the mismatch measure and its split interactions. LPM (OD FE) absorbs origin, destination and day; LPM (region FE) uses region effects plus distance and market-size controls; logit and probit use month and day-of-week effects. Empty _× columns under probit reflect perfect separation of the gap measure.

Table 11

Conversion: Pr(book) over all searches — bus bookings.

	(1) Pooled	(2) Pooled _×	(3) Split	(4) Split _×
<i>LPM (OD FE)</i>				
Connectivity (pooled)	+0.0192 (0.0220)	+0.0192 (0.0219)		
bus component			+0.0009 (0.0237)	-0.0003 (0.0242)
car component			+0.0171 (0.0272)	+0.0183 (0.0273)
Underserved		-0.0235 (0.0224)		+0.0091 (0.0263)
Underserved × conn.		+0.0029 (0.0034)		
Underserved × conn. bus				+0.0112* (0.0067)
Underserved × conn. car				-0.0093 (0.0075)
Log direct supply	+0.0022** (0.0009)	+0.0022** (0.0009)	+0.0022*** (0.0009)	+0.0022** (0.0009)
Price	-0.0007*** (0.0002)	-0.0007*** (0.0002)	-0.0007*** (0.0002)	-0.0006*** (0.0002)
<i>N</i>	20,559,924	20,559,924	20,559,924	20,559,924
<i>Logit</i>				
Connectivity (pooled)	+0.1366 (0.1912)	+0.1233 (0.1892)		
bus component			+0.8653*** (0.2697)	+0.8411*** (0.2687)
car component			-0.6785** (0.2652)	-0.6655** (0.2648)
Underserved		-3.7123 (4.6600)		+0.5495 (2.0710)
Underserved × conn.		+0.4025 (0.6922)		
Underserved × conn. bus				+1.4597** (0.5720)
Underserved × conn. car				-1.2220** (0.5737)
Log direct supply	+0.1097*** (0.0331)	+0.1081*** (0.0331)	+0.1350*** (0.0340)	+0.1330*** (0.0339)
Price	-0.0230*** (0.0059)	-0.0230*** (0.0059)	-0.0245*** (0.0065)	-0.0245*** (0.0064)
<i>N</i>	3,000,000	3,000,000	3,000,000	3,000,000
<i>Probit</i>				
Connectivity (pooled)	+0.0605 (0.0862)	+0.0548 (0.0856)		
bus component			+0.3908*** (0.1228)	+0.3807*** (0.1229)
car component			-0.3063** (0.1278)	-0.3012** (0.1279)
Underserved		-1.3635 (1.7996)		+0.0738 (0.9405)
Underserved × conn.		+0.1386 (0.2689)		
Underserved × conn. bus				+0.4929** (0.2247)
Underserved × conn. car				-0.4082* (0.2438)
Log direct supply	+0.0497*** (0.0149)	+0.0489*** (0.0149)	+0.0604*** (0.0159)	+0.0594*** (0.0159)
Price	-0.0101*** (0.0026)	-0.0101*** (0.0026)	-0.0107*** (0.0029)	-0.0107*** (0.0029)
<i>N</i>	3,000,000	3,000,000	3,000,000	3,000,000

Cluster-robust standard errors (origin) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Columns: (1) connectivity only; (2) adds the mismatch measure (underserved for conversion, gap for mismatch) and its interaction with pooled connectivity; (3) connectivity split into bus and car components; (4) split plus the mismatch measure and its split interactions. LPM (OD FE) absorbs origin, destination and day; LPM (region FE) uses region effects plus distance and market-size controls; logit and probit use month and day-of-week effects. Empty _× columns under probit reflect perfect separation of the gap measure.

B.2 Any mismatch: $\Pr(\text{mismatch} \mid \text{booked})$

Table 12
Any mismatch: Pr(mismatch | booked) — all bookings.

	(1) Pooled	(2) Pooled _×	(3) Split	(4) Split _×
<i>LPM (OD FE)</i>				
Connectivity (pooled)	+0.1086 (0.1349)	+0.1156 (0.1255)		
bus component			-0.1131 (0.0995)	-0.0890 (0.0883)
car component			+0.1760 (0.1587)	+0.1649 (0.1497)
Gap (km)		+0.0054*** (0.0007)		+0.0044*** (0.0008)
Gap × conn.		-0.0005*** (0.0001)		
Gap × conn. bus				-0.0003** (0.0001)
Gap × conn. car				-0.0001 (0.0002)
Log direct supply	-0.0167** (0.0076)	+0.0030 (0.0068)	-0.0171** (0.0072)	+0.0028 (0.0063)
Price	+0.0027* (0.0014)	-0.0005 (0.0013)	+0.0028** (0.0012)	-0.0004 (0.0011)
Bus (booked mode)	-0.3440*** (0.0463)	-0.3209*** (0.0462)	-0.3440*** (0.0464)	-0.3211*** (0.0463)
N	1,301,447	1,301,447	1,301,447	1,301,447
<i>LPM (region FE)</i>				
Connectivity (pooled)	-0.0932 (0.0765)	-0.0904 (0.0753)		
bus component			+0.2494** (0.1071)	+0.2682*** (0.1006)
car component			-0.2792** (0.1213)	-0.2932** (0.1158)
Gap (km)		+0.0052*** (0.0008)		+0.0042*** (0.0008)
Gap × conn.		-0.0005*** (0.0001)		
Gap × conn. bus				-0.0004*** (0.0001)
Gap × conn. car				-0.0001 (0.0002)
Log direct supply	-0.0211*** (0.0066)	-0.0036 (0.0056)	-0.0162** (0.0066)	+0.0018 (0.0059)
Price	+0.0039** (0.0020)	+0.0010 (0.0017)	+0.0031* (0.0019)	+0.0001 (0.0015)
Bus (booked mode)	-0.3275*** (0.0496)	-0.3045*** (0.0486)	-0.3340*** (0.0487)	-0.3114*** (0.0477)
Log market size	-0.0354 (0.0252)	-0.0224 (0.0217)	-0.0508** (0.0241)	-0.0389* (0.0203)
N	1,301,447	1,301,447	1,301,447	1,301,447
<i>Logit</i>				
Connectivity (pooled)	+0.2232 (0.1733)	+0.7543*** (0.2804)		
bus component			+0.1815 (0.2370)	+0.1976 (0.4019)
car component			+0.0493 (0.2401)	+0.5466 (0.3943)
Gap (km)		+19.9489*** (1.8069)		+18.4690*** (2.1133)
Gap × conn.		-1.5238*** (0.2134)		
Gap × conn. bus				-0.4579 (0.3549)
Gap × conn. car				-1.0240** (0.4555)
Log direct supply	-0.1017*** (0.0322)	-0.2502** (0.1023)	-0.0985*** (0.0327)	-0.2480** (0.1016)
Price	+0.0123 (0.0139)	+0.0010 (0.0133)	+0.0116 (0.0139)	+0.0002 (0.0136)
Bus (booked mode)	-1.7190*** (0.3016)	-13.2354*** (0.7173)	-1.7485*** (0.2869)	-13.2301*** (0.7183)
N	1,301,447	1,301,447	1,301,447	1,301,447
<i>Probit</i>				
Connectivity (pooled)	+0.1314 (0.1127)			
bus component			+0.1196 (0.1449)	
car component			+0.0182 (0.1429)	
Log direct supply	-0.0614*** (0.0192)		-0.0594*** (0.0196)	
Price	+0.0073 (0.0085)		+0.0068 (0.0085)	
Bus (booked mode)	-1.0354*** (0.1670)		-1.0564*** (0.1575)	
N	1,301,447		1,301,447	

Cluster-robust standard errors (origin) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Columns: (1) connectivity only; (2) adds the mismatch measure (underserved for conversion, gap for mismatch) and its interaction with pooled connectivity; (3) connectivity split into bus and car components; (4) split plus the mismatch measure and its split interactions. LPM (OD FE) absorbs origin, destination and day; LPM (region FE) uses region effects plus distance and market-size controls; logit and probit use month and day-of-week effects. Empty $\times$ columns under probit reflect perfect separation of the gap measure.

Table 13

Any mismatch: Pr(mismatch | booked) — carpool bookings.

	(1) Pooled	(2) Pooled _×	(3) Split	(4) Split _×
<i>LPM (OD FE)</i>				
Connectivity (pooled)	-0.1052* (0.0600)	-0.0777 (0.0562)		
bus component			+0.0292 (0.0559)	+0.0592 (0.0561)
car component			-0.1150 (0.0728)	-0.1092 (0.0706)
Gap (km)		+0.0057*** (0.0008)		+0.0045*** (0.0009)
Gap × conn.		-0.0006*** (0.0001)		
Gap × conn. bus				-0.0004* (0.0002)
Gap × conn. car				-0.0001 (0.0002)
Log direct supply	-0.0385*** (0.0055)	-0.0135*** (0.0049)	-0.0386*** (0.0054)	-0.0136*** (0.0045)
Price	+0.0035*** (0.0013)	+0.0010 (0.0007)	+0.0035*** (0.0012)	+0.0008 (0.0008)
N	668,425	668,425	668,425	668,425
<i>LPM (region FE)</i>				
Connectivity (pooled)	+0.0173 (0.0334)	+0.0247 (0.0327)		
bus component			+0.1086** (0.0513)	+0.1312*** (0.0493)
car component			-0.0577 (0.0510)	-0.0689 (0.0477)
Gap (km)		+0.0056*** (0.0007)		+0.0045*** (0.0009)
Gap × conn.		-0.0005*** (0.0001)		
Gap × conn. bus				-0.0004* (0.0002)
Gap × conn. car				-0.0001 (0.0002)
Log direct supply	-0.0328*** (0.0066)	-0.0105* (0.0057)	-0.0310*** (0.0066)	-0.0081 (0.0057)
Price	+0.0052*** (0.0013)	+0.0027*** (0.0007)	+0.0045*** (0.0012)	+0.0018** (0.0008)
Log market size	+0.0102 (0.0158)	+0.0123 (0.0170)	-0.0016 (0.0166)	-0.0015 (0.0182)
N	668,425	668,425	668,425	668,425
<i>Logit</i>				
Connectivity (pooled)	+0.2100*** (0.0780)	+0.7543*** (0.2804)		
bus component			+0.0283 (0.2464)	+0.1976 (0.4019)
car component			+0.1829 (0.2984)	+0.5466 (0.3943)
Gap (km)		+18.8500*** (1.0744)		+18.1979*** (1.8843)
Gap × conn.		-1.3275*** (0.1450)		
Gap × conn. bus				-0.1831 (0.5401)
Gap × conn. car				-1.1296* (0.5789)
Log direct supply	-0.0613 (0.0487)	-0.2502** (0.1023)	-0.0611 (0.0489)	-0.2480** (0.1016)
Price	+0.0224*** (0.0060)	+0.0010 (0.0133)	+0.0222*** (0.0049)	+0.0002 (0.0136)
N	668,425	668,425	668,425	668,425
<i>Probit</i>				
Connectivity (pooled)	+0.1315*** (0.0487)			
bus component			+0.0184 (0.1535)	
car component			+0.1139 (0.1858)	
Log direct supply	-0.0382 (0.0302)		-0.0381 (0.0304)	
Price	+0.0138*** (0.0037)		+0.0137*** (0.0030)	
N	668,425		668,425	

Cluster-robust standard errors (origin) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Columns: (1) connectivity only; (2) adds the mismatch measure (underserved for conversion, gap for mismatch) and its interaction with pooled connectivity; (3) connectivity split into bus and car components; (4) split plus the mismatch measure and its split interactions. LPM (OD FE) absorbs origin, destination and day; LPM (region FE) uses region effects plus distance and market-size controls; logit and probit use month and day-of-week effects. Empty $\times$ columns under probit reflect perfect separation of the gap measure.

Table 14

Any mismatch: Pr(mismatch | booked) — bus bookings.

	(1) Pooled	(2) Pooled _×	(3) Split	(4) Split _×
<i>LPM (OD FE)</i>				
Connectivity (pooled)	+0.0357 (0.0560)	+0.0345 (0.0269)		
bus component			-0.0230 (0.0659)	-0.0075 (0.0369)
car component			+0.0544 (0.0363)	+0.0374 (0.0282)
Gap (km)		+0.0050*** (0.0007)		+0.0039*** (0.0009)
Gap × conn.		-0.0004*** (0.0001)		
Gap × conn. bus				-0.0005*** (0.0002)
Gap × conn. car				+0.0001 (0.0002)
Log direct supply	-0.0217*** (0.0025)	-0.0055*** (0.0016)	-0.0216*** (0.0025)	-0.0054*** (0.0016)
Price	+0.0037*** (0.0007)	+0.0007* (0.0003)	+0.0037*** (0.0007)	+0.0007* (0.0003)
N	633,022	633,022	633,022	633,022
<i>LPM (region FE)</i>				
Connectivity (pooled)	-0.3011 (0.2026)	-0.3056 (0.1927)		
bus component			+0.5862*** (0.1791)	+0.6043*** (0.1795)
car component			-0.8122*** (0.2382)	-0.8356*** (0.2366)
Gap (km)		+0.0048*** (0.0010)		+0.0034*** (0.0011)
Gap × conn.		-0.0004*** (0.0001)		
Gap × conn. bus				-0.0006*** (0.0002)
Gap × conn. car				+0.0002 (0.0002)
Log direct supply	-0.0135* (0.0079)	-0.0007 (0.0070)	-0.0125** (0.0050)	+0.0004 (0.0042)
Price	+0.0025 (0.0022)	-0.0001 (0.0019)	+0.0021 (0.0018)	-0.0005 (0.0015)
Log market size	-0.1273*** (0.0435)	-0.0945** (0.0373)	-0.1259*** (0.0363)	-0.0931*** (0.0299)
N	633,022	633,022	633,022	633,022
<i>Logit</i>				
Connectivity (pooled)	+0.1579 (0.5002)	+0.0368 (0.2739)		
bus component			+0.5022 (0.7983)	+0.2245 (0.6142)
car component			-0.2438 (0.5380)	-0.1663 (0.5513)
Gap (km)		+20.3973*** (0.4721)		+20.0209*** (0.3607)
Gap × conn.		-1.6685*** (0.0564)		
Gap × conn. bus				-0.0650 (0.1016)
Gap × conn. car				-1.5955*** (0.0805)
Log direct supply	-0.1400 (0.1125)	-0.6963*** (0.0531)	-0.1318 (0.1187)	-0.6763*** (0.0558)
Price	+0.0039 (0.0257)	+0.0362*** (0.0038)	+0.0033 (0.0256)	+0.0362*** (0.0039)
N	633,022	633,022	633,022	633,022
<i>Probit</i>				
Connectivity (pooled)	+0.0865 (0.2777)			
bus component			+0.2673 (0.4416)	
car component			-0.1309 (0.2959)	
Log direct supply	-0.0796 (0.0596)		-0.0756 (0.0633)	
Price	+0.0024 (0.0148)		+0.0022 (0.0147)	
N	633,022		633,022	

Cluster-robust standard errors (origin) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Columns: (1) connectivity only; (2) adds the mismatch measure (underserved for conversion, gap for mismatch) and its interaction with pooled connectivity; (3) connectivity split into bus and car components; (4) split plus the mismatch measure and its split interactions. LPM (OD FE) absorbs origin, destination and day; LPM (region FE) uses region effects plus distance and market-size controls; logit and probit use month and day-of-week effects. Empty $\times$ columns under probit reflect perfect separation of the gap measure.

B.3 Severe mismatch (vs. matched bookings)

Table 15
Severe mismatch (vs. matched bookings) — all bookings.

	(1) Pooled	(2) Pooled _×	(3) Split	(4) Split _×
<i>LPM (OD FE)</i>				
Connectivity (pooled)	+0.0089 (0.0522)	+0.0207 (0.0318)		
bus component			-0.0082 (0.0507)	+0.0295 (0.0301)
car component			+0.0151 (0.0516)	-0.0041 (0.0336)
Gap (km)		+0.0071*** (0.0007)		+0.0055*** (0.0007)
Gap × conn.		-0.0007*** (0.0001)		
Gap × conn. bus				-0.0005*** (0.0001)
Gap × conn. car				-0.0001 (0.0002)
Log direct supply	-0.0394*** (0.0041)	-0.0141*** (0.0025)	-0.0394*** (0.0041)	-0.0138*** (0.0024)
Price	+0.0054*** (0.0010)	+0.0015*** (0.0004)	+0.0054*** (0.0009)	+0.0014*** (0.0004)
Bus (booked mode)	-0.2703*** (0.0177)	-0.2251*** (0.0151)	-0.2702*** (0.0177)	-0.2251*** (0.0150)
N	1,071,245	1,071,245	1,071,245	1,071,245
<i>LPM (region FE)</i>				
Connectivity (pooled)	+0.0328 (0.0279)	+0.0482** (0.0221)		
bus component			+0.0721* (0.0379)	+0.0946*** (0.0218)
car component			-0.0247 (0.0389)	-0.0322 (0.0206)
Gap (km)		+0.0070*** (0.0007)		+0.0055*** (0.0007)
Gap × conn.		-0.0007*** (0.0001)		
Gap × conn. bus				-0.0005*** (0.0001)
Gap × conn. car				-0.0001 (0.0002)
Log direct supply	-0.0366*** (0.0037)	-0.0139*** (0.0027)	-0.0354*** (0.0036)	-0.0123*** (0.0025)
Price	+0.0059*** (0.0012)	+0.0021*** (0.0006)	+0.0057*** (0.0011)	+0.0018*** (0.0005)
Bus (booked mode)	-0.2713*** (0.0175)	-0.2268*** (0.0149)	-0.2716*** (0.0174)	-0.2270*** (0.0145)
Log market size	-0.0214 (0.0168)	-0.0027 (0.0101)	-0.0253 (0.0169)	-0.0079 (0.0100)
N	1,071,245	1,071,245	1,071,245	1,071,245
<i>Logit</i>				
Connectivity (pooled)	+0.2549*** (0.0655)	+2.2499*** (0.3988)		
bus component			+0.1436 (0.1448)	+0.7666*** (0.2944)
car component			+0.1100 (0.1549)	+1.3894*** (0.2976)
Gap (km)		+3.8644*** (0.2592)		+3.7421*** (0.1965)
Gap × conn.		-0.1571*** (0.0405)		
Gap × conn. bus				-0.0294 (0.0430)
Gap × conn. car				-0.1222*** (0.0383)
Log direct supply	-0.2407*** (0.0222)	+0.0273 (0.0383)	-0.2398*** (0.0218)	+0.0307 (0.0402)
Price	+0.0389*** (0.0057)	-0.0290*** (0.0047)	+0.0385*** (0.0056)	-0.0285*** (0.0047)
Bus (booked mode)	-1.7940*** (0.0862)	-5.1001*** (0.4322)	-1.8086*** (0.0851)	-5.1270*** (0.4154)
N	1,071,245	1,071,245	1,071,245	1,071,245
<i>Probit</i>				
Connectivity (pooled)	+0.1305*** (0.0388)			
bus component			+0.0851 (0.0803)	
car component			+0.0464 (0.0809)	
Log direct supply	-0.1384*** (0.0122)		-0.1378*** (0.0120)	
Price	+0.0230*** (0.0033)		+0.0228*** (0.0033)	
Bus (booked mode)	-1.0055*** (0.0497)		-1.0156*** (0.0498)	
N	1,071,245		1,071,245	

Cluster-robust standard errors (origin) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Columns: (1) connectivity only; (2) adds the mismatch measure (underserved for conversion, gap for mismatch) and its interaction with pooled connectivity; (3) connectivity split into bus and car components; (4) split plus the mismatch measure and its split interactions. LPM (OD FE) absorbs origin, destination and day; LPM (region FE) uses region effects plus distance and market-size controls; logit and probit use month and day-of-week effects. Empty $\times$ columns under probit reflect perfect separation of the gap measure.

Table 16
Severe mismatch (vs. matched bookings) — carpool bookings.

	(1) Pooled	(2) Pooled _x	(3) Split	(4) Split _x
<i>LPM (OD FE)</i>				
Connectivity (pooled)	-0.0732* (0.0412)	-0.0213 (0.0209)		
bus component			+0.0192 (0.0380)	+0.0674** (0.0290)
car component			-0.0795 (0.0499)	-0.0666** (0.0259)
Gap (km)		+0.0079*** (0.0007)		+0.0058*** (0.0009)
Gap × conn.		-0.0008*** (0.0001)		
Gap × conn. bus				-0.0007*** (0.0002)
Gap × conn. car				-0.0000 (0.0002)
Log direct supply	-0.0720*** (0.0052)	-0.0321*** (0.0043)	-0.0721*** (0.0050)	-0.0323*** (0.0038)
Price	+0.0065*** (0.0018)	+0.0023*** (0.0005)	+0.0064*** (0.0019)	+0.0020*** (0.0005)
N	493,453	493,453	493,453	493,453
<i>LPM (region FE)</i>				
Connectivity (pooled)	+0.0621** (0.0259)	+0.0799*** (0.0260)		
bus component			+0.0884** (0.0353)	+0.1228*** (0.0236)
car component			-0.0003 (0.0272)	-0.0133 (0.0185)
Gap (km)		+0.0078*** (0.0007)		+0.0058*** (0.0009)
Gap × conn.		-0.0008*** (0.0001)		
Gap × conn. bus				-0.0007*** (0.0002)
Gap × conn. car				-0.0001 (0.0002)
Log direct supply	-0.0661*** (0.0050)	-0.0302*** (0.0040)	-0.0645*** (0.0051)	-0.0279*** (0.0038)
Price	+0.0076*** (0.0016)	+0.0034*** (0.0007)	+0.0071*** (0.0015)	+0.0027*** (0.0005)
Log market size	+0.0100 (0.0124)	+0.0125 (0.0103)	+0.0012 (0.0116)	+0.0006 (0.0103)
N	493,453	493,453	493,453	493,453
<i>Logit</i>				
Connectivity (pooled)	+0.3931*** (0.0688)	+1.4305*** (0.3205)		
bus component			+0.1912 (0.1746)	+0.6079* (0.3268)
car component			+0.1927 (0.2261)	+0.7559*** (0.2607)
Gap (km)		+3.6212*** (0.2310)		+3.5819*** (0.1773)
Gap × conn.		-0.1117*** (0.0362)		
Gap × conn. bus				-0.0053 (0.0470)
Gap × conn. car				-0.1043** (0.0419)
Log direct supply	-0.2063*** (0.0311)	-0.0844** (0.0397)	-0.2042*** (0.0310)	-0.0745* (0.0410)
Price	+0.0387*** (0.0053)	+0.0039 (0.0048)	+0.0378*** (0.0047)	+0.0038 (0.0048)
N	493,453	493,453	493,453	493,453
<i>Probit</i>				
Connectivity (pooled)	+0.2343*** (0.0414)			
bus component			+0.1146 (0.1023)	
car component			+0.1155 (0.1316)	
Log direct supply	-0.1252*** (0.0192)		-0.1239*** (0.0192)	
Price	+0.0240*** (0.0032)		+0.0235*** (0.0028)	
N	493,453		493,453	

Cluster-robust standard errors (origin) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Columns: (1) connectivity only; (2) adds the mismatch measure (underserved for conversion, gap for mismatch) and its interaction with pooled connectivity; (3) connectivity split into bus and car components; (4) split plus the mismatch measure and its split interactions. LPM (OD FE) absorbs origin, destination and day; LPM (region FE) uses region effects plus distance and market-size controls; logit and probit use month and day-of-week effects. Empty $\times$ columns under probit reflect perfect separation of the gap measure.

Table 17
Severe mismatch (vs. matched bookings) — bus bookings.

	(1) Pooled	(2) Pooled _x	(3) Split	(4) Split _x
<i>LPM (OD FE)</i>				
Connectivity (pooled)	+0.0193 (0.0628)	+0.0062 (0.0317)		
bus component			-0.0415 (0.0604)	-0.0166 (0.0252)
car component			+0.0538 (0.0344)	+0.0173 (0.0222)
Gap (km)		+0.0051*** (0.0007)		+0.0041*** (0.0009)
Gap × conn.		-0.0004*** (0.0001)		
Gap × conn. bus				-0.0004*** (0.0001)
Gap × conn. car				+0.0000 (0.0002)
Log direct supply	-0.0255*** (0.0026)	-0.0094*** (0.0012)	-0.0254*** (0.0026)	-0.0093*** (0.0011)
Price	+0.0040*** (0.0007)	+0.0011*** (0.0002)	+0.0040*** (0.0007)	+0.0011*** (0.0002)
N	577,792	577,792	577,792	577,792
<i>LPM (region FE)</i>				
Connectivity (pooled)	-0.0652 (0.0716)	-0.0422 (0.0486)		
bus component			+0.1183* (0.0715)	+0.0911 (0.0560)
car component			-0.1906* (0.1055)	-0.1400 (0.0912)
Gap (km)		+0.0051*** (0.0007)		+0.0042*** (0.0008)
Gap × conn.		-0.0004*** (0.0001)		
Gap × conn. bus				-0.0004*** (0.0001)
Gap × conn. car				-0.0000 (0.0002)
Log direct supply	-0.0229*** (0.0036)	-0.0088*** (0.0027)	-0.0232*** (0.0032)	-0.0090*** (0.0025)
Price	+0.0045*** (0.0010)	+0.0015** (0.0006)	+0.0044*** (0.0009)	+0.0015*** (0.0006)
Log market size	-0.0820*** (0.0256)	-0.0393** (0.0156)	-0.0828*** (0.0251)	-0.0400*** (0.0154)
N	577,792	577,792	577,792	577,792
<i>Logit</i>				
Connectivity (pooled)	-0.1400 (0.2007)	+4.1225*** (0.6799)		
bus component			-0.3768 (0.3829)	+6.7299*** (1.0021)
car component			+0.1735 (0.3452)	-3.1366*** (0.7249)
Gap (km)		+2.1678*** (0.2914)		+2.8277*** (0.2443)
Gap × conn.		+0.0432 (0.0436)		
Gap × conn. bus				+0.4157*** (0.0711)
Gap × conn. car				-0.3412*** (0.0665)
Log direct supply	-0.3381*** (0.0334)	+0.1636 (0.1133)	-0.3419*** (0.0385)	+0.0352 (0.1079)
Price	+0.0420*** (0.0080)	-0.0420*** (0.0159)	+0.0426*** (0.0085)	-0.0241* (0.0144)
N	577,792	577,792	577,792	577,792
<i>Probit</i>				
Connectivity (pooled)	-0.0717 (0.1006)			
bus component			-0.1984 (0.1893)	
car component			+0.0945 (0.1638)	
Log direct supply	-0.1746*** (0.0176)		-0.1773*** (0.0205)	
Price	+0.0234*** (0.0042)		+0.0236*** (0.0045)	
N	577,792		577,792	

Cluster-robust standard errors (origin) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Columns: (1) connectivity only; (2) adds the mismatch measure (underserved for conversion, gap for mismatch) and its interaction with pooled connectivity; (3) connectivity split into bus and car components; (4) split plus the mismatch measure and its split interactions. LPM (OD FE) absorbs origin, destination and day; LPM (region FE) uses region effects plus distance and market-size controls; logit and probit use month and day-of-week effects. Empty $\times$ columns under probit reflect perfect separation of the gap measure.

B.4 Connectivity dose-response (LPM within OD and day)

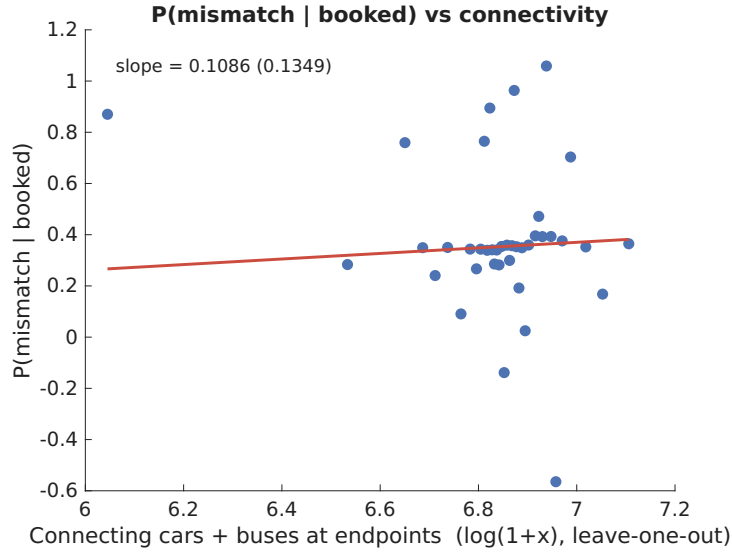


Figure 11

Probability that a booking is a mismatch vs. endpoint connectivity, within OD and day (binned partial-residual plot).

Note: Binned partial-residual plot of the probability that a booking is a near-match (booked OD $\neq$ searched OD) against endpoint connectivity, among bookings. Endpoint connectivity is $\log(1 + \text{cars} + \text{buses})$ arriving at and departing from the searched origin and destination, excluding the searched route (leave-one-out). The outcome and connectivity are residualized on origin, destination and day fixed effects and on the controls (displayed price and log direct supply); points are means within connectivity ventiles (20 bins) and the line is the fitted linear-probability slope, reported in the panel. Full sample of bookings.

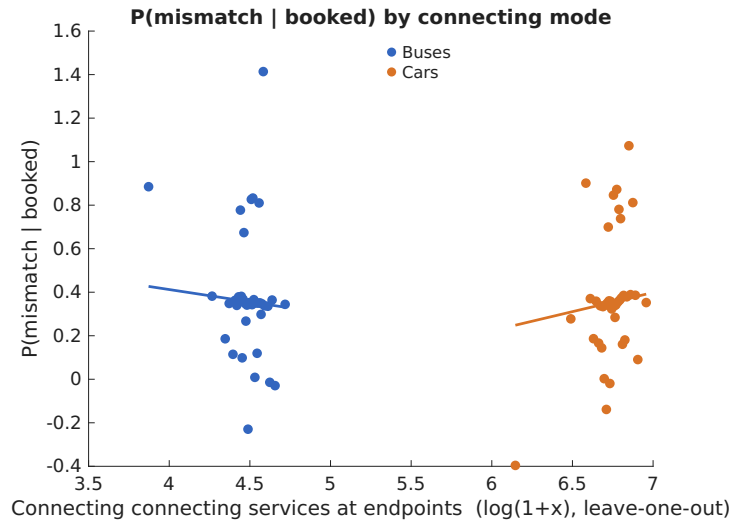


Figure 12

The same dose-response separating connecting buses from connecting cars.

Note: As in Figure 11, but endpoint connectivity is split into its bus and carpooling components, entered simultaneously. Points are binned partial residuals and the lines are the fitted linear-probability slopes for connecting buses and connecting cars, within origin, destination and day fixed effects and the same controls. Full sample of bookings.

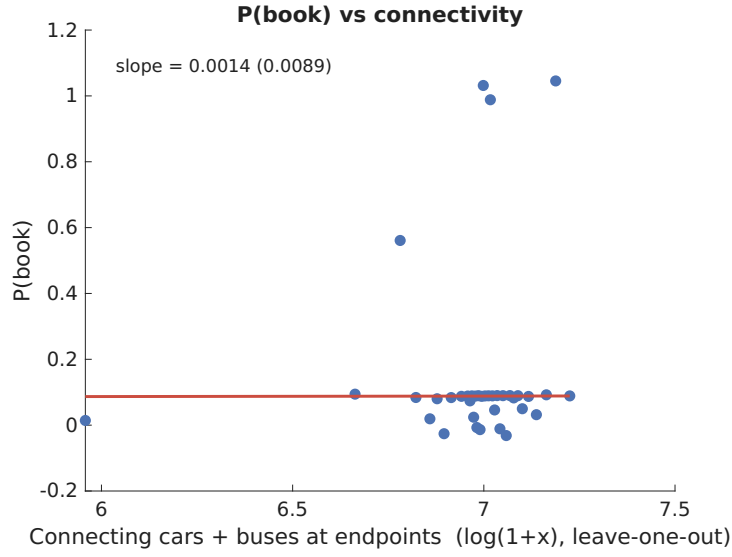


Figure 13

Conversion probability (whether a search books) vs. endpoint connectivity, within OD and day.

Note: Binned partial-residual plot of the conversion probability (whether a search results in a booking) against endpoint connectivity, over all searches. Connectivity, the fixed effects (origin, destination, day) and the controls are defined as in Figure 11; points are means within connectivity ventiles (20 bins) and the line is the fitted linear-probability slope. Full sample of searches.

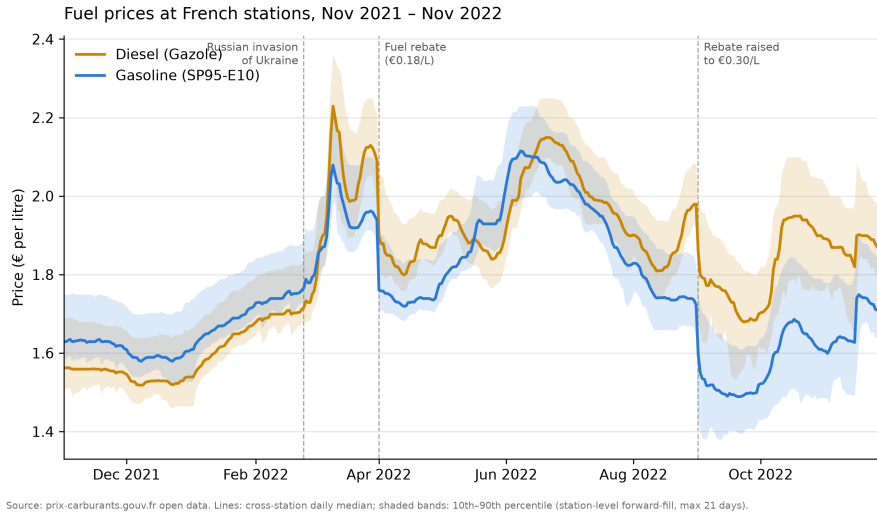


Figure 14

Fuel prices in France, 2021–2022

Note: Daily fuel prices in France over the sample period (November 2021–November 2022): the average station price of diesel and of gasoline (E10). These series, matched to each trip's origin municipality and departure date, form the fuel-cost component of the variable cost-per-seat instrument used to address price endogeneity (Section 5.2).

C Additional Results and Robustness checks

C.1 Price elasticities

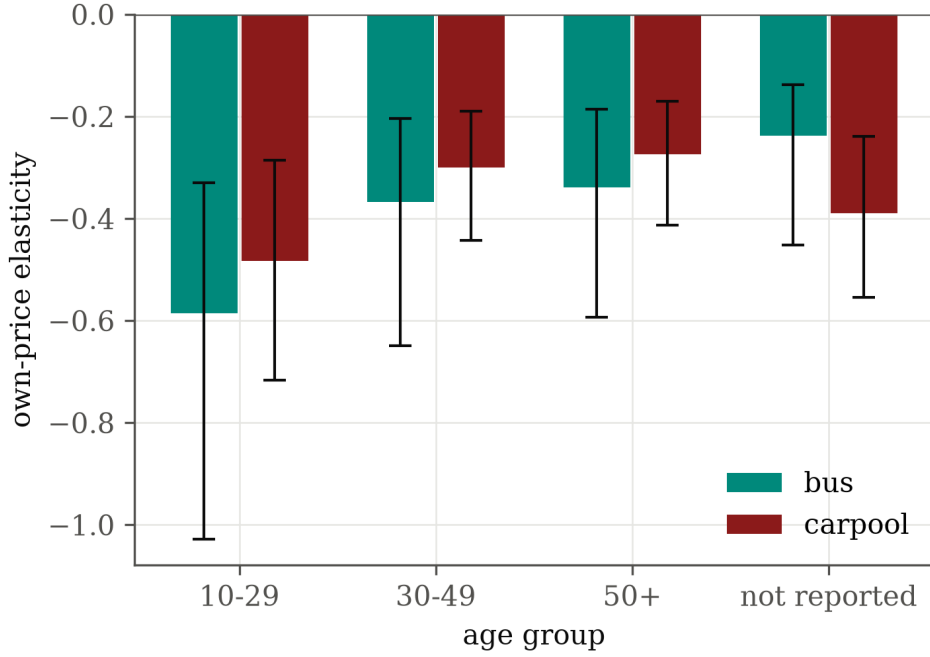
Table 18
PRICE ELASTICITIES AND DIVERSION RATIOS

	Bus	Carpool	Outside option
<i>Panel A: Price elasticities</i>			
Bus	−1.095	0.170	
Carpool	0.154	−1.059	
Outside option	0.047	0.056	
<i>Panel B: Diversion ratios</i>			
Bus	–	0.167	0.833
Carpool	0.134	–	0.866

Notes: Panel A: the entry in row j and column k is the percentage change of the aggregate share of j when the price of every cell (departure window) of mode k rises by 1%. The diagonal holds the own-price elasticities, the off-diagonal entries the cross-price elasticities, and the outside-option row the change in the share of searchers who do not book; the outside option has no price, so its column is empty. Panel B: the entry in row j and column k is the diversion ratio $D_{jk} = -(\partial s_k / \partial p_j) / (\partial s_j / \partial p_j)$, the share of the demand mode j loses after an increase in its price that goes to k ; each row sums to one. Both panels are mode-level aggregates implied by the stage-1 estimates (column (3) of Table 7) and the 2SLS price coefficient of column (6) of Table 8 (−0.036; gap bands and network variables, month, weekday, route and departure-window fixed effects, instruments IV11). Each searcher’s price coefficient is that coefficient plus the stage-1 price $\times$ demographics interactions. Sample: 89 routes, 1,458 markets; the market size is the number of searches. Observed aggregate shares: bus 0.046, carpool 0.055, outside option (searched, not booked) 0.899. Point estimates; standard errors not computed.

The estimated own-price elasticities are −1.095 for buses and −1.059 for carpooling (Table 18). Two considerations guide their interpretation. First, these are mode-level aggregates: each entry is the response of a mode’s total share to a uniform change in the price of all of its cells. Drivers and the platform, however, set prices at the ride (departure-window) level, and the own-price elasticity of an individual cell is mechanically larger than that of the mode as a whole, since a price increase on one cell diverts demand to the remaining cells of the same mode as well as to the other mode and to the outside option. The mode-level numbers are the relevant object for the aggregate counterfactuals that follow, but they understate the price sensitivity faced by a single ride. Second, our magnitudes are in line with the scarce evidence on long-distance travel. They are close, in particular, to [Astier et al. \(2023\)](#), the nearest study using individual-level BlaBlaCar data, who finds that a 1% increase in the price of all carpool rides lowers the carpooling share by about 1.2%, against 1.06% in our estimates; the small remaining difference is consistent with carpooling and buses being weak substitutes in our data, as most of the demand a mode loses after a price increase accrues to the outside option rather than to the rival mode (Panel B).

Figure 15
Own-price elasticity by Age Group



Notes: Booker-level own-price elasticity of each mode. Bars show the median across bookers in each age group whereas whiskers show the interquartile range.

Two features of our setting may nonetheless push these estimates toward zero. The price coefficient is identified from within-route variation: with route (OD), month, weekday and departure-window fixed effects, α is recovered from the residual movement of prices within a route, and although the excluded instruments are relevant ($F = 29$), this variation is limited, so some attenuation cannot be ruled out. In addition, prices enter as displayed averages over the cells of a market, which introduces classical measurement error and biases the coefficient further toward zero. This also means that we do not take into account the intertemporal dimension of the preference of consumers for prices. Notably, consumers booking many days in advance are likely to be more price sensitive than consumers booking just the day before the trip, who are more likely to care less about the price and more about other factors for booking a trip. We therefore read the estimated elasticities as, if anything, a lower bound on the true price sensitivity; correspondingly, the break-even surcharge of the next section is best interpreted as an upper bound on how much the platform could charge before harming consumers on average.

C.2 Robustness

Table 19
ROBUSTNESS: STAGE TWO ESTIMATES WITH STAGE-ONE SPECIFICATION (2)

	OLS		2SLS (IV11)			
	(1)	(2)	(3)	(4)	(5)	(6)
Price $\times$ bus	−0.012 (0.001)		−0.030 (0.006)			
Price $\times$ carpool	−0.011 (0.001)		−0.035 (0.006)			
Price		−0.012 (0.001)		−0.025 (0.005)	−0.032 (0.005)	−0.053 (0.007)
Gap sum (100 km)	−0.315 (0.010)	−0.315 (0.010)	−0.322 (0.012)	−0.317 (0.011)		
Origin gap, minor					−0.063 (0.060)	0.084 (0.068)
Origin gap, severe					−1.172 (0.051)	−0.861 (0.079)
Destination gap, minor					−0.016 (0.058)	0.108 (0.069)
Destination gap, severe					−1.198 (0.050)	−0.763 (0.084)
Log searches						0.015 (0.090)
Log offers						0.359 (0.033)
Log results displayed						0.501 (0.031)
Departure 0–6h	−1.091 (0.027)	−1.090 (0.027)	−1.093 (0.027)	−1.092 (0.027)	−1.103 (0.027)	−1.188 (0.031)
Departure 12–18h	0.333 (0.025)	0.332 (0.025)	0.312 (0.026)	0.320 (0.026)	0.310 (0.026)	0.281 (0.027)
Departure 18–24h	−0.627 (0.024)	−0.628 (0.024)	−0.603 (0.026)	−0.608 (0.026)	−0.619 (0.026)	−0.672 (0.030)
Bus	0.489 (0.038)	0.462 (0.019)	0.370 (0.045)	0.486 (0.022)	0.409 (0.026)	2.003 (0.064)
Constant	−4.595 (0.263)	−4.584 (0.261)	−4.297 (0.237)	−4.425 (0.246)	−3.995 (0.197)	−6.974 (0.532)
Observations	9,533	9,533	9,533	9,533	9,533	8,630
F-stat			70.4 / 72.8	29.1	29.2	18.9

Notes: Stacked regressions of the cell mean utilities $\hat{\delta}$ from stage 1 (column (2) of Table 7). Cell: searched route $\times$ date $\times$ mode $\times$ six-hour departure window; market: route $\times$ date. Sample: 89 routes, 1,458 markets; observations are the cells with complete data (fewer with network variables, which are missing for some markets). Price: transaction price in euros (carpool: driver amount plus commission and VAT); Price $\times$ bus and Price $\times$ carpool are mode-specific coefficients. Gaps: distance between the searched and the booked origin (destination). Gap sum: mean origin plus destination gap of the cell's bookings, in 100 km. Gap bands: shares of the cell's bookings with a minor (0.5–25 km) or a severe (>25 km) gap; base: no gap. Network variables, in $\ln(1+x)$ and treated as exogenous: searches on the route-date, offers of the mode for the searched origin–destination, results displayed for the mode. Departure dummies relative to 6–12h; Bus: bus dummy (carpool base). Fixed effects: month, weekday, route (departure window reported). Instruments, interacted with the mode dummy: IV11 = trip cost per seat and per seat-km (fuel, driver time and tolls over seats; bus on diesel, carpool on E10 prices). F-stat: heteroskedasticity-robust first-stage F on the excluded instruments; with two prices, Sanderson–Windmeijer conditional F for the bus / carpool price (not robust), because the unconditional F is mechanically inflated: each mode's price is zero on the other mode's cells, which the mode-interacted instruments fit exactly. Heteroskedasticity-robust standard errors in parentheses.

Table 20
ROBUSTNESS: STAGE TWO ESTIMATES, 50-ROUTE SAMPLE

	OLS		2SLS (IV11)			
	(1)	(2)	(3)	(4)	(5)	(6)
Price $\times$ bus	−0.009 (0.001)		−0.025 (0.004)			
Price $\times$ carpool	−0.005 (0.001)		−0.021 (0.004)			
Price		−0.008 (0.001)		−0.023 (0.004)	−0.024 (0.004)	−0.028 (0.005)
Gap sum (100 km)	−0.202 (0.007)	−0.203 (0.007)	−0.209 (0.010)	−0.210 (0.010)		
Origin gap, minor					−0.001 (0.049)	0.024 (0.049)
Origin gap, severe					−0.733 (0.039)	−0.536 (0.049)
Destination gap, minor					0.001 (0.040)	0.071 (0.047)
Destination gap, severe					−0.731 (0.042)	−0.538 (0.052)
Log searches						−0.111 (0.056)
Log offers						0.258 (0.020)
Log results displayed						0.247 (0.019)
Departure 0–6h	−0.960 (0.019)	−0.960 (0.019)	−0.955 (0.020)	−0.955 (0.020)	−0.962 (0.020)	−0.984 (0.021)
Departure 12–18h	0.090 (0.017)	0.089 (0.017)	0.078 (0.017)	0.078 (0.017)	0.073 (0.017)	0.056 (0.017)
Departure 18–24h	−0.513 (0.016)	−0.518 (0.016)	−0.484 (0.019)	−0.490 (0.018)	−0.506 (0.018)	−0.535 (0.019)
Bus	0.105 (0.027)	−0.009 (0.013)	0.107 (0.034)	0.006 (0.014)	−0.050 (0.019)	0.787 (0.039)
Constant	−3.207 (0.079)	−3.111 (0.077)	−2.772 (0.141)	−2.718 (0.139)	−2.477 (0.129)	−2.933 (0.218)
Observations	8,059	8,059	8,059	8,059	8,059	7,579
F-stat			66.0 / 69.1	36.1	35.5	24.5

Notes: Stacked regressions of the cell mean utilities $\hat{\delta}$ from stage 1 (column (3) of Table 7). Cell: searched route $\times$ date $\times$ mode $\times$ six-hour departure window; market: route $\times$ date. Sample: 50 routes, 1,200 markets; observations are the cells with complete data (fewer with network variables, which are missing for some markets). Price: transaction price in euros (carpool: driver amount plus commission and VAT); Price $\times$ bus and Price $\times$ carpool are mode-specific coefficients. Gaps: distance between the searched and the booked origin (destination). Gap sum: mean origin plus destination gap of the cell's bookings, in 100 km. Gap bands: shares of the cell's bookings with a minor (0.5–25 km) or a severe (>25 km) gap; base: no gap. Network variables, in $\ln(1+x)$ and treated as exogenous: searches on the route-date, offers of the mode for the searched origin–destination, results displayed for the mode. Departure dummies relative to 6–12h; Bus: bus dummy (carpool base). Fixed effects: month, weekday, route (departure window reported). Instruments, interacted with the mode dummy: IV11 = trip cost per seat and per seat-km (fuel, driver time and tolls over seats; bus on diesel, carpool on E10 prices). F-stat: heteroskedasticity-robust first-stage F on the excluded instruments; with two prices, Sanderson–Windmeijer conditional F for the bus / carpool price (not robust), because the unconditional F is mechanically inflated: each mode's price is zero on the other mode's cells, which the mode-interacted instruments fit exactly. Heteroskedasticity-robust standard errors in parentheses.

Table 21
ROBUSTNESS: STAGE TWO ESTIMATES WITH DIFFERENT INSTRUMENTAL VARIABLES

	OLS	2SLS (IV13)	2SLS (IV14)	2SLS (IV11)
	(1)	(2)	(3)	(4)
Price	−0.008 (0.001)	−0.015 (0.002)	−0.008 (0.002)	−0.018 (0.004)
Gap sum (100 km)	−0.198 (0.007)	−0.199 (0.007)	−0.198 (0.007)	−0.199 (0.008)
Departure 0–6h	−0.931 (0.019)	−0.932 (0.019)	−0.931 (0.019)	−0.932 (0.019)
Departure 12–18h	0.105 (0.016)	0.098 (0.016)	0.105 (0.016)	0.094 (0.017)
Departure 18–24h	−0.512 (0.015)	−0.501 (0.016)	−0.513 (0.015)	−0.496 (0.017)
Bus	−0.023 (0.013)	−0.009 (0.013)	−0.024 (0.013)	−0.003 (0.014)
Constant	−3.176 (0.175)	−3.088 (0.163)	−3.184 (0.178)	−3.046 (0.162)
Observations	9,533	9,533	9,533	9,533
F-stat		91.4	95.0	29.1
Sargan p -value		0.000	0.000	0.000

Notes: Stacked regressions of the cell mean utilities $\hat{\delta}$ from stage 1 (column (3) of Table 7). Cell: searched route $\times$ date $\times$ mode $\times$ six-hour departure window; market: route $\times$ date. Sample: 89 routes, 1,458 markets; observations are the cells with complete data. Price: transaction price in euros (carpool: driver amount plus commission and VAT). Gaps: distance between the searched and the booked origin (destination). Gap sum: mean origin plus destination gap of the cell's bookings, in 100 km. Departure dummies relative to 6–12h; Bus: bus dummy (carpool base). Fixed effects: month, weekday, route (departure window reported). Instruments, interacted with the mode dummy: IV13 = trip cost per seat, per seat-km and a Hausman price (mean price of the same mode and departure window on the same date on routes sharing no endpoint); IV14 = fuel cost, fuel cost $\times$ distance and the Hausman price; IV11 = trip cost per seat and per seat-km (fuel, driver time and tolls over seats; bus on diesel, carpool on E10 prices). F-stat: heteroskedasticity-robust first-stage F on the excluded instruments. Sargan p -value: test of the overidentifying restrictions. Heteroskedasticity-robust standard errors in parentheses.

Table 22
ROBUSTNESS: STAGE TWO ESTIMATES WITH DIFFERENT GAP VARIABLES

	2SLS (IV11)		
	(1)	(2)	(3)
Price	−0.018 (0.004)	−0.028 (0.004)	−0.023 (0.004)
Gap sum (100 km)	−0.199 (0.008)		
Origin gap (any)		−0.432 (0.031)	
Destination gap (any)		−0.411 (0.030)	
Origin gap, minor			−0.052 (0.040)
Origin gap, severe			−0.732 (0.035)
Destination gap, minor			−0.004 (0.039)
Destination gap, severe			−0.749 (0.035)
Departure 0–6h	−0.932 (0.019)	−0.940 (0.020)	−0.939 (0.019)
Departure 12–18h	0.094 (0.017)	0.090 (0.018)	0.088 (0.017)
Departure 18–24h	−0.496 (0.017)	−0.491 (0.018)	−0.502 (0.017)
Bus	−0.003 (0.014)	−0.166 (0.018)	−0.053 (0.018)
Constant	−3.046 (0.162)	−2.681 (0.162)	−2.775 (0.133)
Observations	9,533	9,533	9,533
F-stat	29.1	29.2	29.2

Notes: Stacked regressions of the cell mean utilities $\hat{\delta}$ from stage 1 (column (3) of Table 7). Cell: searched route $\times$ date $\times$ mode $\times$ six-hour departure window; market: route $\times$ date. Sample: 89 routes, 1,458 markets; observations are the cells with complete data. Price: transaction price in euros (carpool: driver amount plus commission and VAT). Gaps: distance between the searched and the booked origin (destination). Gap sum: mean origin plus destination gap of the cell's bookings, in 100 km. Gap (any): share of the cell's bookings with a gap above 0.5 km. Gap bands: shares of the cell's bookings with a minor (0.5–25 km) or a severe (>25 km) gap; base: no gap. Departure dummies relative to 6–12h; Bus: bus dummy (carpool base). Fixed effects: month, weekday, route (departure window reported). Instruments, interacted with the mode dummy: IV11 = trip cost per seat and per seat-km (fuel, driver time and tolls over seats; bus on diesel, carpool on E10 prices). F-stat: heteroskedasticity-robust first-stage F on the excluded instruments. Heteroskedasticity-robust standard errors in parentheses.

Table 23
ROBUSTNESS: STAGE TWO ESTIMATES WITH DIFFERENT NETWORK VARIABLES

	2SLS (IV11)			
	(1)	(2)	(3)	(4)
Price	−0.019 (0.004)	−0.032 (0.004)	−0.020 (0.004)	−0.036 (0.005)
Origin gap, minor	0.077 (0.043)	−0.067 (0.040)	0.071 (0.043)	0.035 (0.047)
Origin gap, severe	−0.564 (0.039)	−0.659 (0.041)	−0.307 (0.048)	−0.544 (0.053)
Destination gap, minor	0.078 (0.044)	0.024 (0.038)	0.066 (0.044)	0.064 (0.047)
Destination gap, severe	−0.471 (0.042)	−0.630 (0.041)	−0.196 (0.052)	−0.498 (0.058)
Log searches	−0.182 (0.047)	−0.023 (0.047)	−0.180 (0.047)	−0.009 (0.060)
Log offers	0.385 (0.016)		0.402 (0.016)	0.211 (0.022)
Log results displayed		0.370 (0.015)		0.299 (0.021)
Log offers × gap sum			−0.045 (0.005)	
Departure 0–6h	−0.990 (0.019)	−0.958 (0.021)	−0.983 (0.019)	−0.996 (0.022)
Departure 12–18h	0.088 (0.016)	0.078 (0.017)	0.087 (0.016)	0.071 (0.018)
Departure 18–24h	−0.555 (0.017)	−0.506 (0.018)	−0.547 (0.017)	−0.539 (0.020)
Bus	0.435 (0.025)	0.749 (0.038)	0.451 (0.025)	0.892 (0.043)
Constant	−2.924 (0.287)	−3.740 (0.293)	−3.013 (0.297)	−4.419 (0.361)
Observations	8,630	9,533	8,630	8,630
F-stat	19.4	27.3	19.0	18.9

Notes: Stacked regressions of the cell mean utilities $\hat{\delta}$ from stage 1 (column (3) of Table 7). Cell: searched route × date × mode × six-hour departure window; market: route × date. Sample: 89 routes, 1,458 markets; observations are the cells with complete data (fewer with network variables, which are missing for some markets). Price: transaction price in euros (carpool: driver amount plus commission and VAT). Gaps: distance between the searched and the booked origin (destination). Gap bands: shares of the cell’s bookings with a minor (0.5–25 km) or a severe (>25 km) gap; base: no gap. Network variables, in $\ln(1 + x)$ and treated as exogenous: searches on the route-date, offers of the mode for the searched origin–destination, results displayed for the mode; offers are also interacted with the gap sum. Departure dummies relative to 6–12h; Bus: bus dummy (carpool base). Fixed effects: month, weekday, route (departure window reported). Instruments, interacted with the mode dummy: IV11 = trip cost per seat and per seat-km (fuel, driver time and tolls over seats; bus on diesel, carpool on E10 prices). F-stat: heteroskedasticity-robust first-stage F on the excluded instruments. Heteroskedasticity-robust standard errors in parentheses.

Table 24
ROBUSTNESS: STAGE TWO ESTIMATES WITH DIFFERENT FIXED EFFECTS

	2SLS (IV11)			
	(1)	(2)	(3)	(4)
Price	−0.004 (0.002)	−0.005 (0.002)	−0.006 (0.002)	−0.018 (0.004)
Gap sum (100 km)	−0.203 (0.007)	−0.186 (0.007)	−0.187 (0.007)	−0.199 (0.008)
Distance (100 km)	−0.108 (0.015)	−0.104 (0.015)	−0.096 (0.016)	
Departure 0–6h		−0.945 (0.020)	−0.940 (0.020)	−0.932 (0.019)
Departure 12–18h		0.107 (0.017)	0.106 (0.017)	0.094 (0.017)
Departure 18–24h		−0.522 (0.016)	−0.520 (0.016)	−0.496 (0.017)
Bus	−0.030 (0.016)	−0.036 (0.013)	−0.032 (0.013)	−0.003 (0.014)
Constant	−3.562 (0.034)	−3.330 (0.029)	−3.355 (0.032)	−3.046 (0.162)
Month FE	✓	✓	✓	✓
Time window FE	×	✓	✓	✓
Weekday FE	×	×	✓	✓
Route FE	×	×	×	✓
Observations	9,533	9,533	9,533	9,533
F-stat	60.3	60.3	59.2	29.1

Notes: Stacked regressions of the cell mean utilities $\hat{\delta}$ from stage 1 (column (3) of Table 7). Cell: searched route $\times$ date $\times$ mode $\times$ six-hour departure window; market: route $\times$ date. Sample: 89 routes, 1,458 markets; observations are the cells with complete data. Price: transaction price in euros (carpool: driver amount plus commission and VAT). Gaps: distance between the searched and the booked origin (destination). Gap sum: mean origin plus destination gap of the cell's bookings, in 100 km. Departure dummies relative to 6–12h; Bus: bus dummy (carpool base). Distance is included when route fixed effects are not. Fixed effects: as indicated. Instruments, interacted with the mode dummy: IV11 = trip cost per seat and per seat-km (fuel, driver time and tolls over seats; bus on diesel, carpool on E10 prices). F-stat: heteroskedasticity-robust first-stage F on the excluded instruments. Heteroskedasticity-robust standard errors in parentheses.