PASSENGER DEMAND FOR UKRZALIZNYTSIA: DRIVERS AND PERSPECTIVES FOR DYNAMIC PRICING

by

Dmytro Hushchin

A thesis submitted in partial fulfillment of the requirements for the degree of

MA in Business and Financial Economics

Kyiv School of Economics

2025

Thesis Supervisor: Professor Oleh Nivievskyi

Approved by

Head of the KSE Defense Committee, Professor [Type surname, name]

Date ___________________________________

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

UZ Ukrzaliznytsia, the Ukrainian Railways

EDA Exploratory Data Analysis

OLS Ordinary Least Squares

FE Fixed Effects

HC3 Heteroskedasticity-Consistent standard errors, type 3

IQR Interquartile Range

CHAPTER 1. INTRODUCTION

Ukrzaliznytsia (UZ) is a national railway operator that provides both passenger and freight services in Ukraine. This is one of the largest railways in Europe, as the company transported over 65 million passengers in 2024 (Ministry for Communities, Territories and Infrastructure Development of Ukraine, 2024). Since the full-scale Russian invasion in 2022, the importance of railway transport has grown even more: civil aviation is suspended, international mobility is restricted, and millions of people have relocated to central and western regions. This has created excess demand, while supply has remained limited. At the same time, railways have become the primary type of international transport.

Despite its scale, UZ remains unprofitable. In 2024, the passenger segment produced an operating loss of UAH 18.1 billion (Interfax-Ukraine, 2024). UZ ended the year with a net loss of UAH 2.7 billion (Interfax-Ukraine, 2024). Because the passenger segment is loss-making, UZ cross-subsidizes it from freight operations. In 2024, the freight segment generated an operating profit of about UAH 20.4 billion (Interfax-Ukraine, 2024). However, the company has proposed raising freight tariffs by 37% (GMK Center, 2024). Such a significant increase raises logistics costs for exporters. The European Business Association highlights that losses should be covered through direct budget support (EBA, 2025). In 2025, the government already allocated UAH 4.3 billion for partial compensation (EBA, 2025). In October 2025, the Cabinet of Ministers provided UAH 8 billion of financial assistance to UZ (Cabinet of Ministers of Ukraine, 2025). This results in a burden for taxpayers, especially during wartime when resources are urgently needed for defense.

The main problem is that UZ's passenger segment remains loss-making. The regulated pricing model may be one reason and a solution. The research question is: "What are the drivers of passenger demand for Ukrzaliznytsia, and what are the perspectives for dynamic pricing?"

To answer this question, I collected a dataset of ticket availability for 36 routes within Ukraine using a custom-built web scraper. Based on a literature review and a descriptive analysis, I formulated four hypotheses. The first hypothesis checks whether the booking curve is non-linear and shows a U-shaped pattern. The second one tests how routes, classes, and calendar factors explain final occupancy. The third hypothesis checks whether high-demand routes have lower volatility. The fourth one tests the stability of the booking curve for each route and class. All assumptions were tested using regression models and nonparametric tests.

The results show that final occupancy is mainly explained by route and class, rather than time and day of the week. Both exploratory data analysis and regression results show that passenger demand follows a U-shaped pattern. It means higher booking activity at the beginning and toward the end of the sales period. Similar, though weaker, U-shaped temporal effects have been observed in airline markets (Piga et al., 2017). However, late bookings within the last two days account for less than 25% of total demand. In comparison, Italian high-speed rail data show that about 75% of tickets are purchased in the final week and 46% in the last 2 days before departure (Hetrakul & Cirillo, 2014). The potential impact of last-minute dynamic pricing for UZ is limited.

Given these findings, UZ could consider raising fares, introducing early-access premium options, extending the booking horizon, and optimizing wagon allocation.

CHAPTER 2. INDUSTRY OVERVIEW AND RELATED STUDIES

2.1. Industry overview

Ukrzaliznytsia is a state-owned natural monopoly that controls all passenger and freight rail transport in Ukraine (EBRD, 2023). However, this situation is not unique compared to Europe. Many European railway operators, such as Deutsche Bahn in Germany and SNCF in France, also remain state-owned (Lethbridge, 2019). However, unlike UZ, they operate within liberalized markets. In Europe, private and foreign operators compete for services and network access. "By liberalising operations, competition between operators is possible, allowing users to choose between different operators and prices… To achieve liberalisation, a complete separation between the rail infrastructure manager and the passenger and freight operators is needed, as a minimum" (European Commission, 2023). The United Kingdom is the main exception. Since the 1990s, it has privatized most of its rail services. The infrastructure, however, remains publicly owned through Network Rail (Lethbridge, 2019).

Passenger transport in Ukraine is concentrated around major cities. In 2018, Kyiv, Kharkiv, Dnipro, Lviv, and Odesa together generated around 57% of total passenger rail travel (European Commission, 2023). Kyiv alone accounts for over 25% of all demand (European Commission, 2023). Passenger kilometres reached approximately 25 billion, of which 91% were domestic trips, including suburban, regional, and long-distance routes. The long-distance segment consists of several train types: Intercity (IC), Intercity+ (IC+), night express and night fast services, and regional trains. Night fast trains account for 69% of total long-distance passenger kilometres. Passenger demand is seasonal. For instance, August volumes in 2018 were 42% higher than in February, showing a summer travel peak (European Commission, 2023).

The government, not the market, sets Ukrzaliznytsia's ticket prices. "The entire rail network is controlled by the Cabinet of Ministers of Ukraine, with the Ministry of

Infrastructure functioning as the main executive body controlling Ukraine's transportation infrastructure (roads, trains, and communications)" (European Commission, 2023).

The Ministry of Infrastructure sets tariff tables that link price to distance: each range of kilometres (e.g., 1-10 km, 11-20 km, and so on) has a fixed fare for each train and coach type (Ministry of Infrastructure of Ukraine, 2018). The total ticket price combines a base fare and an additional seat or berth charge. The extra fare also increases incrementally with distance. These rates are then adjusted by official coefficients based on the season, day of travel, and train category (e.g., Intercity, express, or regional). The Ministry also sets a fixed cost for transporting empty coaches – 6.69 UAH per coach-kilometre (Ministry of Infrastructure of Ukraine, 2018).

The Law on Rail Transport and the Law on Prices and Price Formation define rail transport as a public utility obliged to serve social and economic needs (European Commission, 2023). The Antimonopoly Committee of Ukraine has confirmed that UZ holds a 100% dominant position in the domestic rail market. As a result, both passenger and freight tariffs remain state-regulated and cross-subsidized. Passenger services are subsidized at the expense of freight operations. This structure limits commercial flexibility and profitability (European Commission, 2023).

Financial results support that freight transport forms the backbone of the Ukrainian railway sector (Interfax-Ukraine, 2024). Bulk commodities dominate the freight market. In 2019, UZ transported 68.3 million tonnes of iron and manganese ore, 40 million tonnes of coal, and 39.8 million tonnes of wheat. Despite the technical advantages of rail for heavy cargo, road transport has become competitive on short- to medium-distance routes (European Commission, 2023). It is especially important for agriculture due to quicker delivery and fewer infrastructure limitations (European Commission, 2023).

While passenger tariffs are strictly regulated, freight pricing shows partial market adjustment. "Rates for shipping in UZ's wagons respond to overall demand… To determine the daily rate, UZ either holds an auction or employs a procedure that considers the trend in demand and current degree of utilization of the rolling stock" (World Bank, 2025, p. 52). These pricing mechanisms help to adapt to demand changes quickly. For instance, UZ's freight rates increased in 2022-2023 due to the launch of the Black Sea Grain Initiative. The introduction of a market-based pricing system shows a shift in the Ukrainian freight sector towards efficiency (World Bank, 2025).

However, the Ukrainian railway sector faces unique challenges. The railway's physical condition remains a main limitation (European Commission, 2023). UZ's rolling stock is heavily aged. The average age is 42 years for electric locomotives, 32 years for mainline diesel locomotives, and 38 years for shunters. More than 600 locomotives require repairs. The average ages of electric and diesel multiple units are 21 and 18 years, respectively, which still exceeds European standards (European Commission, 2023).

Since a full-scale Russian invasion in 2022, UZ's priorities shifted to network security and people evacuation (World Bank, 2025). By November 2024, over 126 stations and 500 km of track were damaged. With international support, UZ restored 17 destroyed bridges and over 80 engineering structures, acquired 18 modular bridges, 90 units of heavy equipment, and 200 new flat wagons to increase export (World Bank, 2025).

Ukraine's railway system operates under complete government control. This model is typical for Central and Eastern Europe. However, the Railway Law aims to move towards European standards by separating infrastructure from operations, removing cross-subsidies, and opening access to private operators, similar to reforms in Poland and the Czech Republic (World Bank, 2025).

Many operators in Europe apply dynamic pricing to increase revenue and manage demand. So, tariff reform becomes a key component of the transition towards European market standards.

2.2. International practices

Dynamic pricing is widely used in airlines and high-speed rail transportation. Since the 1970s, many airline operators have adjusted ticket prices based on booking time and service class. Nowadays, companies integrate advanced tools such as AI-based dynamic pricing. In 2024, Lufthansa Group adopted dynamic pricing, which applies neural networks and Bayesian models to optimize fares (PhocusWire, 2024).

This approach is also becoming popular in rail transportation. China tested dynamic pricing on the busiest railway line between Beijing and Shanghai in December 2020 (Business Traveller, 2020). In May 2024, JR Kyushu became the first Shinkansen operator in Japan to apply lower fares for advance online tickets between Fukuoka and Kumamoto. Passengers pay one of three fares, replacing the previous fixed-price system (Kyodo News, 2024). In India, premium express change fares are based on the real-time availability of berths (Rail Mitra, 2019).

In Europe, the trend is mixed. "International train fares within Europe all seem to have this dynamic pricing where the fare goes up as the date approaches and more tickets are sold" (Price of Travel, 2022). France, Germany, Spain, and the UK use advanced prices for high-speed and intercity trains. Some countries, including Switzerland and the Scandinavian region, still use fixed domestic fares. Nevertheless, they may offer super saver tickets for advance purchase (Price of Travel, 2022).

In the United States, Amtrak uses dynamic pricing across all its routes, with fares rising and falling in real time based on consumer demand, similar to airline revenue management. Passengers may secure the lowest fares even days before departure if demand is low, but prices increase as trains fill up (The Urbanist, 2019).

2.3. Literature overview

Classical research on revenue management combines discrete choice modeling (e.g., multinomial logit or latent class) with fare and seat allocation optimization (Hetrakul & Cirillo, 2011; Vulcano, van Ryzin & Chaar, 2010; Hetrakul & Cirillo, 2014).

Hetrakul and Cirillo (2013) analyzed booking data from Italian high-speed trains to understand the factors that affect when people buy tickets. They segmented passengers by trip length (short, medium, or long) and used discrete choice models. The results showed that ticket price, days to departure, and day of the week influence demand. They also discovered that grouping passengers by when they book (early vs. late) is more insightful than demographics. Some people book early to save money, while others book late despite higher prices. This means that travel behavior and context matter more than the traveler's identity.

In 2014, Hetrakul and Cirillo extended their research by adding a revenue optimization. They used discrete-choice models to simulate how different pricing and seat-allocation strategies affect passenger behavior and revenue. Key insight is that serving more short-haul passengers is more profitable than reserving seats for long-haul travelers. This is because a single seat can be sold multiple times along shorter portions of a route. Their model showed that revenue increased from 3.29% to 20.44% compared to fixed pricing.

Cao et al. (2024) studied China's Shandong high-speed rail network. They found that different routes and train types show different demand patterns. The authors used a space-time service network. For each trip, the model calculates a total cost based on ticket price, travel time, comfort, and convenience. To predict which trains people choose, the authors use a special path-size logit model. They found that using different prices for each route and train type is more profitable.

Manchiraju et al. (2025) expanded earlier work by studying both pricing and capacity decisions together. They used data from Japan's Shinkansen high-speed rail. They developed a unified model to optimize both ticket prices and the number of coaches per train. The results confirmed that railways can earn more and serve more passengers by adjusting both pricing and supply capacity.

All these studies rely on detailed booking data. In this research, we have only aggregated train-day data. That is why it is impossible to use discrete-choice or seat-allocation models. So, the thesis will focus on demand drivers using descriptive analysis, regression models, and booking curves.

CHAPTER 3. DATA

To get the necessary data, I used a custom-built web scraper of the Ukrzaliznytsia ticketing system. Each day, the script collects information on all trains and wagons for selected routes. The dataset covers 36 routes between 11 cities, including Kyiv, Lviv, Odesa, Kharkiv, Dnipro, Poltava, Zaporizhzhia, Zhytomyr, Chernihiv, Kryvyi Rih, and Sumy. These cities are Ukrainian economic and population centers. In 2018, Kyiv, Kharkiv, Dnipro, Lviv, and Odesa together generated around 57% of total passenger rail travel (European Commission, 2023). Kyiv alone accounts for over 25% of all demand (European Commission, 2023). Technical issues also limited the selection of routes. Collecting data is computationally heavy and requires significant cloud resources.

For each train and wagon, the dataset provides:

- parsing_datetime: date and time when the data were scraped;
- departure_datetime, arrival_datetime: scheduled departure and arrival times;
- from_station, to_station, route: origin, destination stations, and named route;
- train_number: identifier of the train service;
- wagon_name: coach identifier within the train;
- class, seat_type: service class and type of seating (bed, seat, etc.);
- price: ticket price for that wagon and class segment;
- seat_count, available_seats: total seats and seats still unsold at the scrape time.

From raw data, I calculate the needed variables. Booked seats are total seats minus available seats. The occupancy rate is the share of booked seats and is calculated as the number of booked seats divided by the total number of seats. Journey duration is the difference between departure and arrival times. Days to departure are the difference between the departure and the parsing datetime. Detailed distance information for each train and route was unavailable, so I calculated the straight-line distance between cities as an approximation.

Table 1 contains descriptive statistics for the key variables. The dataset has 440,853 rows. The average ticket price is 509 UAH. Prices range from 81 UAH to 1251 UAH. The mean number of seats per wagon is 53. Wagons have between 20 and 134 seats. The average price per kilometer is 0.92 UAH.

Table 1 – Descriptive statistics of key variables

variable	count	mean	std	min	25%	median	75%	max
price	440853	508.89	256.72	81	314	482	717	1251
seat_count	440853	53.00	18.22	20	40	54	56	134
available_seats	440853	26.60	22.75	1	8	22	39	131
seats_booked	440853	26.40	16.11	0	13	26	35	133
price_per_km	440853	0.92	0.35	0.24	0.59	0.91	1.16	2

Table 2 demonstrates variation across routes. The average price ranges from 170 UAH (Kyiv-Sumy) to nearly 880 UAH (Lviv-Dnipro). Occupancy ranges from 28% (Kyiv-Dnipro) to 85% (Lviv-Kharkiv). The routes can be divided into three distinct groups, based on the price and occupancy. I did not use clusterization, so the segmentation is descriptive rather than algorithmic.

The first group consists of expensive, low-occupancy routes. For example, Kyiv-Dnipro and Dnipro-Kyiv (600 UAH, 28-35% occupancy) show oversupply. This segment requires either a price decrease or a demand stimulation. The second group includes expensive and high-occupancy routes such as Lviv-Kharkiv (600-630 UAH, 80-85% occupancy) and Lviv-Dnipro (880 UAH, 59-68% occupancy). This segment is the best choice for dynamic pricing. The last group is medium-distance routes such as Lviv-Zaporizhzhia and Kyiv-Sumy (170-230 UAH, 40-55% occupancy). They are closer to the equilibrium of price and occupancy. This group leaves less room for pricing adjustments.

Table 2 – Descriptive statistics by route

route	count	mean price	median price	mean occupancy	median occupancy
Zaporizhzhia → Kyiv	45221	541.58	502	0.53	0.55
Kyiv → Zaporizhzhia	39053	551.57	517	0.49	0.47
Kharkiv → Kyiv	35419	514.96	489	0.49	0.48
Kyiv → Kharkiv	26091	544.86	489	0.42	0.38
Kyiv → Dnipro	21736	595.04	517	0.35	0.28
Lviv → Kharkiv	17307	626.62	613	0.80	0.85
Dnipro → Kyiv	16760	600.77	517	0.35	0.30
Kyiv → Sumy	14216	170.64	171	0.43	0.39
Kharkiv → Lviv	12948	579.73	387	0.77	0.82
Lviv → Zaporizhzhia	12051	232.57	225	0.54	0.50
Lviv → Dnipro	10415	878.21	982	0.59	0.68
Odesa → Kyiv	10414	690.69	791	0.57	0.60
Zaporizhzhia → Lviv	9666	211.72	220	0.50	0.46
Sumy → Kyiv	9567	172.05	171	0.54	0.53
Lviv → Odesa	9382	697.62	791	0.55	0.60
Lviv → Darnytsia	8627	599.05	498	0.43	0.39
Kyiv → Odesa	8026	714.84	791	0.58	0.62
Mukachevo → Odesa	7653	313.73	337	0.73	0.78
Odesa → Lviv	7516	712.48	791	0.59	0.65
Darnytsia → Lviv	7031	603.38	545	0.49	0.45

Table 3 shows variation across classes. The regular coupe has moderate prices and a high occupancy (552 UAH, 70%). It is a core class with strong demand. Platzkart offers low prices and has a high occupancy (232 UAH, 63%). It is the most affordable and widely used option. In contrast, 2nd class (451 UAH, 48%) and especially 1st class (662 UAH, 37%) show lower occupancy. It suggests oversupply or limited willingness to pay. The 3rd class has low prices and minimal occupancy (96 UAH, 12%). Women-only and child-only classes have the highest prices and occupancy (713-727 UAH, 84-89%).

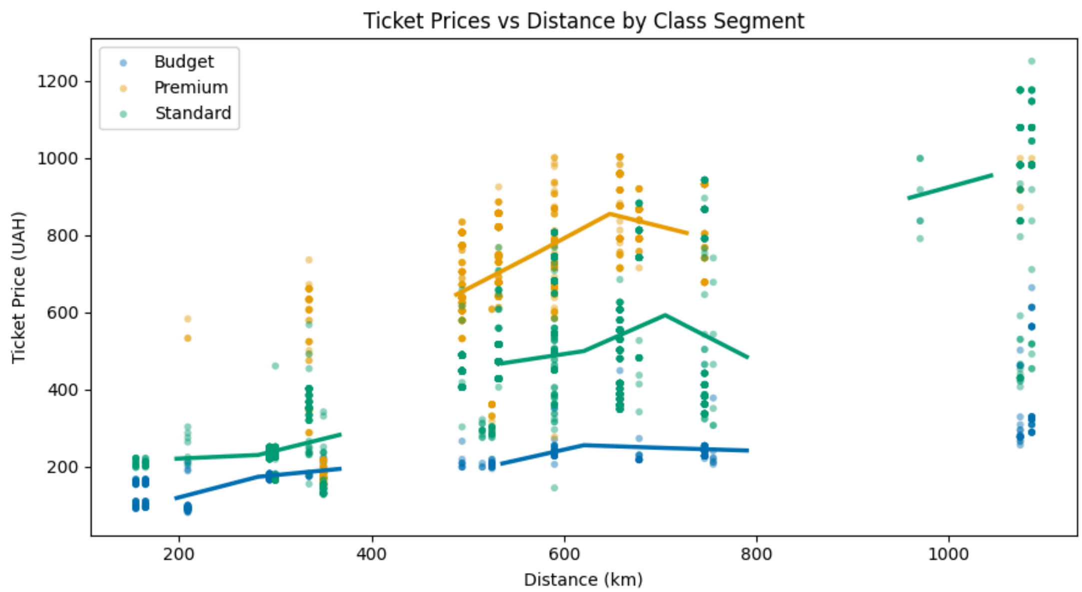
Table 3 – Descriptive statistics by class

route	count	mean price	median price	mean occupancy	median occupancy
Regular coupe	127943	551.77	412	0.70	0.78
2nd class	124119	450.67	472	0.48	0.46
1st class	112596	662.29	727	0.37	0.30
Platzkart	47473	231.68	209	0.63	0.69
3rd class	13500	95.76	95	0.12	0.09
Women coupe	8021	712.73	681	0.84	0.88
Child coupe	7200	726.85	678	0.84	0.89

To analyze demand patterns across different market segments, wagons were grouped into three categories based on price level: premium (1st class, women-only and child-only coupe), standard (2nd class, regular coupe), and budget (platzkart, 3rd class).

Figure 1 shows a scatter plot of ticket prices versus distance. The solid lines represent the average trend for each class. The plot shows that prices increase with distance for all classes. At the same time, the price levels remain consistently separated. Budget fares are the lowest, Standard is in the middle range, and Premium is the highest across all distances.

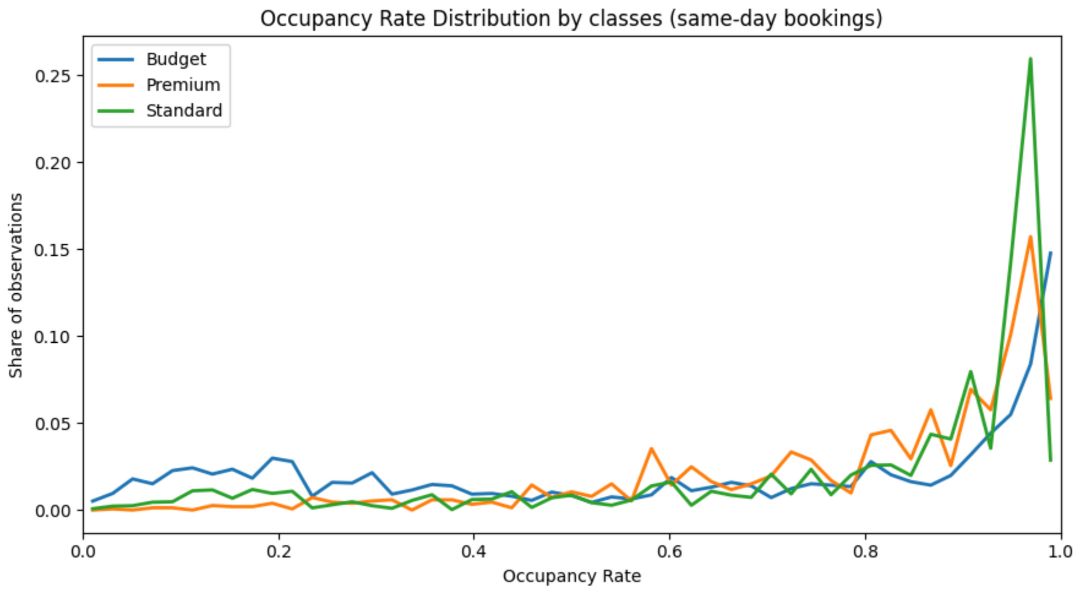
Figure 1 – Prices and distance by class segment



Note: Budget (Platzkart, 3rd class), Standard (Coupe, 2nd class), Premium (Luxury, 1st class, Female/Child coupe)
Source: Author's calculations based on UZ scraped data

Figure 2 illustrates the final occupancy by class segment. Final occupancy means the occupancy on the day of departure. The distribution of final occupancy is skewed towards high values across the premium and standard segments. The budget segment has a similar pattern but flatter distribution. Overall, this indicates that most trains sell out by departure time.

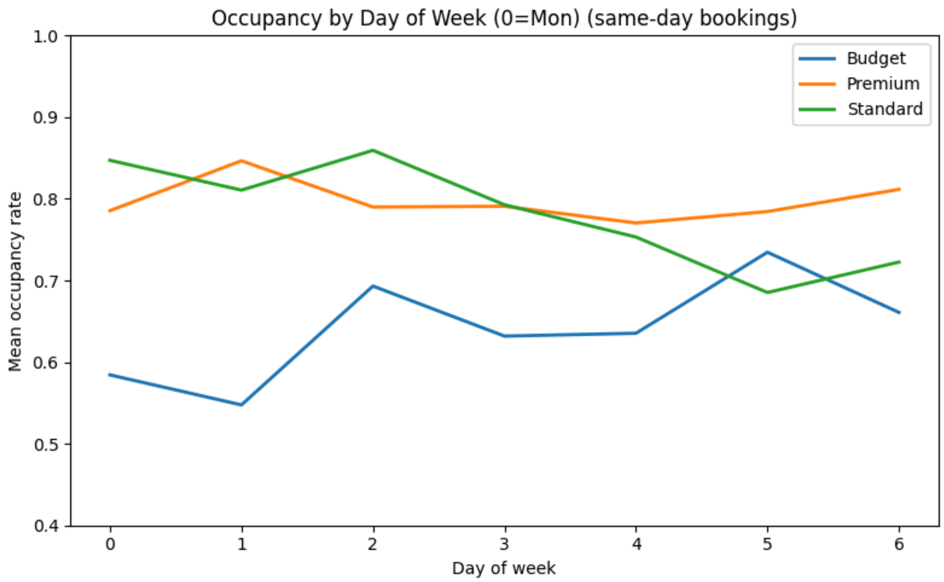
Figure 2 – Final occupancy rate distribution by class segment



Note: Budget (Platzkart, 3rd class), Standard (Coupe, 2nd class), Premium (Luxury, 1st class, Female/Child coupe)
Source: Author's calculations based on UZ scraped data

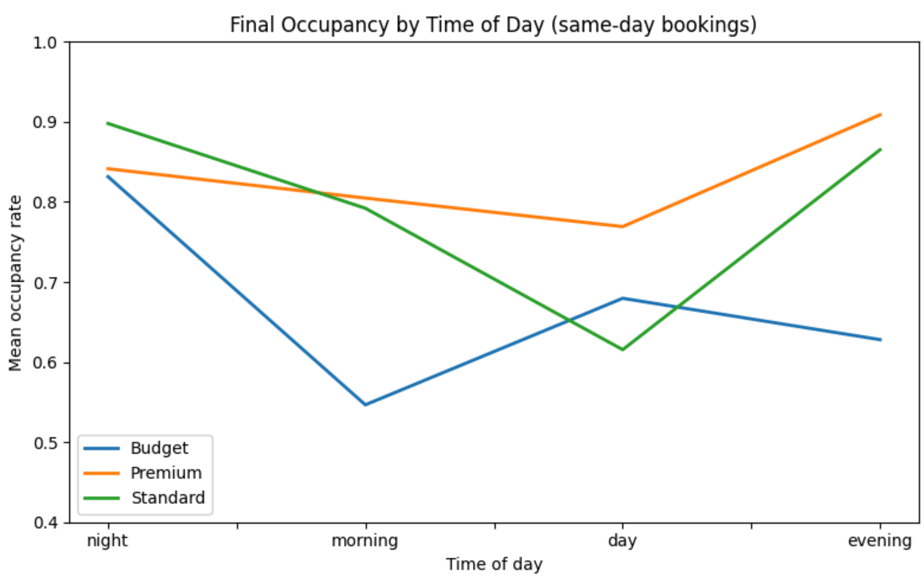
Figures 3 and 4 demonstrate final occupancy by class segment, day of the week, and time of the day. Premium wagons have high occupancy throughout the week and throughout the day (about 80%). Standard wagons show high weekday (80%) but lower weekend occupancy (70%). This segment reaches near-full occupancy (90%) at night but falls during the day (60%). Budget wagons show low occupancy on Monday and Tuesday (55%) and rise on Saturday (70%). They also have lower but stable loads throughout the day (about 65%). The results indicate that passenger demand across all classes shows a mixed pattern. Business trips might be the primary driver of weekday and nighttime travel, while weekend and daytime travel are likely for leisure passengers.

Figure 3 – Final occupancy by day of the week and class segment



Note: Budget (Platzkart, 3rd class), Standard (Coupe, 2nd class), Premium (Luxury, 1st class, Female/Child coupe)
Source: Author's calculations based on UZ scraped data

Figure 4 – Final occupancy by time of the day and class segment

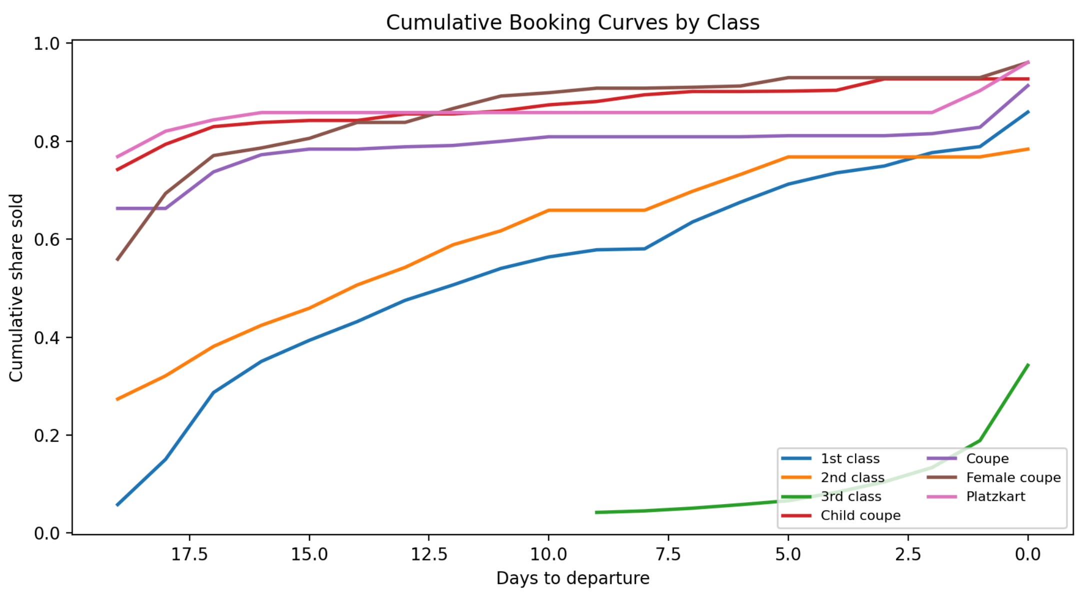


Note: Budget (Platzkart, 3rd class), Standard (Coupe, 2nd class), Premium (Luxury, 1st class, Female/Child coupe)
Source: Author's calculations based on UZ scraped data

Figure 5 illustrates how quickly seats sell by class. On the horizontal axis, the figure shows the number of days to departure, counted backward from the departure date. On the vertical axis, it demonstrates the share of seats booked by that day. The slope of each curve reflects how ticket sales accelerate as departure approaches.

Coupe wagons reach around 80% occupancy when tickets are released. These segments have strong demand and are less affected by price changes. First- and second-class bookings increase gradually. In contrast, Platzkart and third-class trains show slower early sales and a sharper increase on departure day.

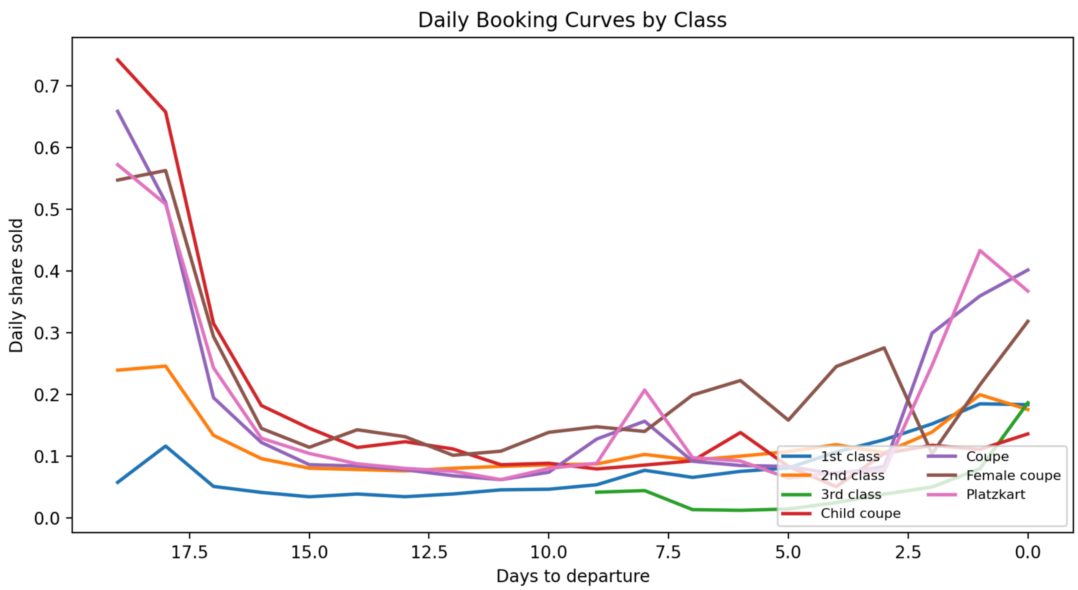
Figure 5 – Cumulative booking curves by class



Source: Author’s calculations based on UZ scraped data

Cumulative curves show an increase in sales at the start and the end of the booking period. This pattern aligns with findings by Hetrakul and Cirillo (2014). Figure 6 shows the daily share of tickets sold for each class, illustrating a U-shaped demand pattern.

Figure 6 – Daily booking curves by class



Source: Author's calculations based on UZ scraped data

CHAPTER 4. METHODOLOGY

I am unable to use discrete-choice models because they require individual-level data. Instead, I focus on regression models and nonparametric tests. Based on a literature review and a descriptive analysis, I formulated four hypotheses. The first hypothesis checks whether the booking curve is non-linear and shows a U-shaped pattern. The second one tests how routes, classes, and calendar factors explain final occupancy. The third hypothesis checks whether high-demand routes have lower volatility. The fourth one tests the stability of the booking curve for each route and class.

4.1. Hypothesis 1

Prior studies show that many passengers are late buyers, which motivates last-minute price increases (Talluri & van Ryzin, 2004; Hetrakul & Cirillo, 2014). However, UZ data indicate another effect – a high level of occupancy on the first day of sales. This combination suggests that demand may follow a U-shaped temporal pattern, with peaks at the beginning and end of the booking window.

To test this relationship, I estimate several panel regression models with entity fixed effects. Each route-class combination represents a panel entity observed repeatedly. Fixed effects serve as a set of intercepts that account for time-invariant differences across routes and classes. It is important because some routes or classes are more popular.

To capture potential nonlinear and proportional patterns, three functional forms are compared:

- Log-log model. Both ticket sales and days to departure are expressed in logarithmic form to capture proportional effects;
- Level-quadratic model. Daily sales depend on both the linear and squared terms of days to departure;
- Log-quadratic model. Ticket sales are log-transformed. Daily sales depend on both the linear and squared terms of days to departure.

The specification of model 1:

$$log(1 + Sold_{i,t}) = \alpha_i + \beta 1 * log(1 + DaysToDept) + \varepsilon_{i,t} \qquad (1)$$

A dependent variable is the logarithm of daily ticket sales for route-class combination i on t days before departure. The logarithm helps smooth out the effect of large ticket sales and reduces heteroskedasticity. The independent variable is the logarithm of the number of days remaining until departure, capturing the elasticity of ticket sales with respect to time remaining. Epsilon is the error term.

The specification of model 2:

$$Sold_{i,t} = \alpha_i + \beta 1 * DaysToDept + \beta 2 * DaysToDept^2 + \varepsilon_{i,t} \qquad (2)$$

A dependent variable is the daily number of tickets sold for route-class combination i on t days before departure. Both a linear and a squared term of days before departure are included to capture potential nonlinear patterns, such as a U-shaped relationship between time to departure and ticket sales intensity. Epsilon is the error term.

The specification of model 3:

$$log(1 + Sold_{i,t}) = \alpha_i + \beta 1 * DaysToDept + \beta 2 * DaysToDept^2 + \varepsilon_{i,t} \qquad (3)$$

A dependent variable is the logarithm of daily ticket sales for route-class combination i on t days before departure. The logarithm helps smooth out the effect of large ticket sales and reduces heteroskedasticity. Both a linear and a squared term of days before departure are included to capture potential nonlinear patterns, such as a U-shaped relationship between time to departure and ticket sales intensity. Epsilon is the error term.

4.2. Hypothesis 2

It is expected that final train occupancy is shaped primarily by structural and service-related factors rather than calendar effects.

Preliminary data indicate that variations across routes and service classes account for most of the differences in occupancy. The time of day and the day of the week have weaker effects. I assume that journey duration also contributes to demand variation. Seasonal patterns could not be evaluated due to limited data. To test these relationships, I restrict the data to the day of departure, when the booking process is complete, and the realized load factor can be observed. The model is estimated using pooled OLS with HC3 heteroskedasticity-robust standard errors. Because the dependent variable (final occupancy rate) is bounded between 0 and 1 and right-skewed, it is logit-transformed to stabilize variance and allow linear estimation.

$$y = log((y + \varepsilon) / (1 - y + \varepsilon)),\ \varepsilon = 10^{-6} \tag{4}$$

$$\begin{aligned} Occupancy_{transformed} &= \beta \ + \beta1 * Weekend_i + \beta2 * Holiday_i \\ \beta3 * JourneyHours_i &+ \gamma * Route(i) + \gamma * Class(i) \\ &+ \gamma * TimeOfDay(i) + \varepsilon_i \end{aligned} \tag{5}$$

The dependent variable is the logit-transformed final occupancy for train-class combination i; weekend and holiday capture calendar effects; journey hours controls for trip length; and the gamma terms are dummies for route, service class, and time of departure.

4.3. Hypothesis 3

It is expected that highly loaded departures have more stable day-to-day demand than low-load departures.

Stable booking curves are easier to forecast. If high-load trips have smoother sales dynamics, operators can adjust prices with greater confidence. Firstly, I compute the final occupancy for each departure and class. Then I split the sample into high-load and low-load groups using quantiles (upper and lower 33%). For each segment, I measure booking volatility as the variance of daily sales over the selling horizon and compare mean variances across groups.

I conduct a Welch's t-test on log-transformed variances and a Mann-Whitney U test as a robustness check. The t-test checks whether the average volatility differs between the two groups. At the same time, the Mann-Whitney test checks whether the distributions vary overall. It ensures that extreme values or distributional assumptions do not drive the result.

4.4. Hypothesis 4

Passenger booking behavior is expected to be consistent within the same route and service class.

If booking curves are similar across segments, the median curve can be used as a benchmark for demand forecasting. To test this, I calculate the interquartile range (IQR) of cumulative booking curves for each class-route combination. Then I apply a one-sample Wilcoxon signed-rank test to check whether the median IQR is significantly less than 20%.

CHAPTER 5. RESULTS

Exploratory data analysis revealed several patterns in passenger behavior. In this chapter, I test these patterns formally using the hypotheses and econometric methods described earlier.

5.1. Hypothesis 1

It is expected that passenger demand follows a U-shaped pattern relative to the departure date.

Table 4 – H1 models comparison

Specification	β_1	β_2	Pivot ($-\beta_1/2\beta_2$)	R^2 (within)	F-statistic
Log-Log	0.0978	–	–	0.002	103.1
Level-Quadratic	-5.1021	0.3104	8.22	0.057	1321.9
Log-Quadratic	-0.1897	0.0112	8.51	0.073	1714.8

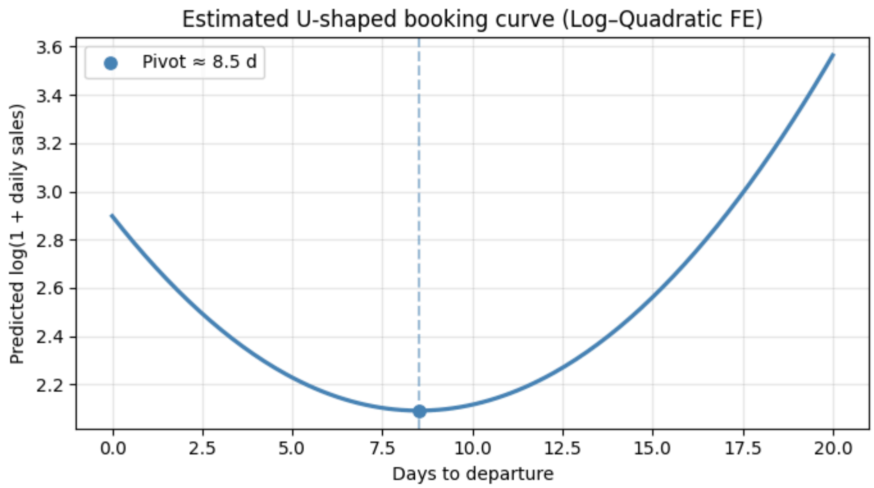
Notes: All specifications estimated with entity fixed effects and clustered SEs by entity. Pivot reported for quadratic specifications only. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

The Model 3 explains 7.3% of within-entity variation in daily ticket sales (R^2 = 7.3%). The F-statistic for overall model significance is 1714.8 ($p < 0.001$). The F-test for poolability ($p < 0.001$) supports the inclusion of fixed effects. There are systematic differences between routes and classes.

The variable days_to_dep is negative (-0.1897, $p < 0.001$), while the squared term is positive (0.0112, $p < 0.001$). This means a nonlinear relationship: ticket sales are initially high, then decline and accelerate again closer to departure. So, passenger demand follows a U-shaped temporal pattern.

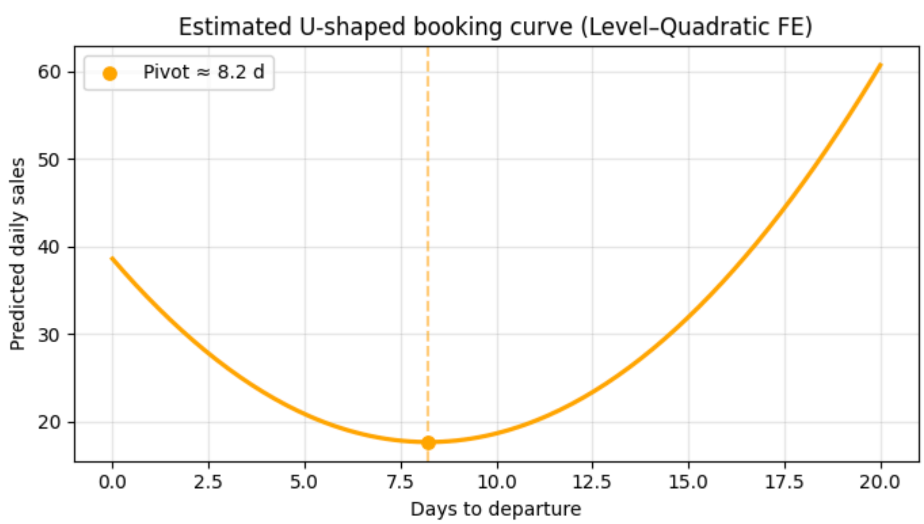
Pivot point marks the lowest level of daily ticket sales within the booking horizon. It separates the early-booking phase (strong initial demand right after tickets become available) from the late-purchase phase (sales during the final week before departure). Visualizing the fitted curve confirms this U-shaped profile (Figure 7).

Figure 7 – Estimated U-shaped booking curve (Log-Quadratic FE)



Even though the log-quadratic model shows a slightly higher R^2, it is harder to interpret because the results are on a logarithmic scale. To make the pattern more transparent and easier to understand, a similar curve was estimated using the level-quadratic model, which shows the same U-shaped trend in actual ticket sales (Figure 8).

Figure 8 – Estimated U-shaped booking curve (Level-Quadratic FE)



5.2. Hypothesis 2

It is expected that final train occupancy is shaped primarily by structural and service-related factors rather than calendar effects.

Table 5 – Determinants of final occupancy (Pooled OLS)

Constant	3.244*** (0.341)
Weekend	0.023 (0.061)
Holiday	-0.109 (0.101)
Journey hours	-0.093*** (0.020)
R2 (adj)	0.394
F-statistic	61.21

Notes: Pooled OLS with HC3 robust SE in parentheses; * p<0.10, ** p<0.05, *** p<0.01. The dependent variable is the logit-transformed final occupancy rate.

The regression explains about 39% of the variation in final occupancy (R^2 = 39%). The results show that calendar and temporal effects (weekends, holidays, time of day) are not statistically significant, which aligns with the earlier descriptive statistics.

Major routes (Lviv-Kharkiv, Odesa-Kyiv, Lviv-Odesa) show significantly higher occupancies. In contrast, routes such as Zhytomyr-Kyiv or Kryvyi Rih-Zaporizhzhia have low utilization. The coefficient on journey hours is negative but small (-0.093, p = 0.001). This means that longer trips have lower final occupancy. Service class also matters. The Coupe and Platzkart classes reach the highest utilization (p < 0.001). Second and third-class trains show consistently lower utilization.

The included factors jointly explain a large share of variation (F-statistic = 61.21, $p < 0.001$). So, final occupancy depends on route and class characteristics, but not on calendar or temporal effects.

5.3. Hypothesis 3

It is expected that highly loaded departures have more stable day-to-day demand.

The difference in daily sales variance between the high-load and low-load groups is about 1.5 times (3024 vs. 2009). Welch's t-test ($t = 3.49$, $p = 0.0005$) and the Mann-Whitney test ($p < 0.001$) confirm that this difference is statistically significant. High-load trains exhibit greater daily sales variance. This suggests that demand for such trains fluctuates more strongly during the sales period. It may limit the predictability of booking behavior and require more cautious implementation of dynamic pricing strategies than on low-load routes. Contrary to expectations, high-load trains show greater variance, indicating that H3 is rejected.

Table 6 – Volatility of daily sales vs. final load

Metric	Value
Mean variance (high-load)	3024
Mean variance (low-load)	2009
Ratio (high/low)	1.5
t-test (Welch) – t, p	3.491, 0.0005
Mann-Whitney – U, p	573854, 0.0000

5.4. Hypothesis 4

The passenger booking pattern is expected to be stable within the same route and service class.

The median IQR width of cumulative sales is only 0.6% across 144 segments. About 92% of segments have a width smaller than 20%. The Wilcoxon signed-rank test rejects the null hypothesis ($p < 0.001$). Results confirm that sales are concentrated in a short window. It means passenger behavior can be predicted with high accuracy.

Table 7 – Stability of booking curves

Metric	**Value**
Tau (threshold)	0.2
Segments	144
Median width	0.006
Share below tau	0.924
Wilcoxon (stat, p, n)	281, 0.0000, 144

CHAPTER 6. CONCLUSIONS AND RECOMMENDATIONS

This thesis studied the main factors of passenger demand for Ukrzaliznytsia (UZ). To collect the data, I built a custom web scraper for parsing UZ's ticketing website. Each day, the script collects information on seat availability for all trains and wagons. The dataset contains more than 400000 observations from 36 routes and 11 cities. The methodology uses exploratory data analysis, regression methods, and nonparametric tests. There are several insights for implementing market-based pricing.

First, passenger demand has a U-shaped booking pattern. Ticket sales are high on release day, decline in the middle period, and increase again closer to departure. However, the share of late bookings is low compared to the literature. This limits the effectiveness of last-minute pricing. Secondly, final occupancy depends mainly on route and class. Journey hours have a small effect. Time and day of the week play no significant role. Third, booking curves are stable for each route-class segment. The results confirm that UZ operates in a market with high and stable demand. However, it is essential to consider the broader context of rail transportation in Ukraine.

Firstly, the war caused a population migration from the eastern and southern regions of the country. It sharply increased passenger flows. Secondly, international mobility is restricted. Most men are unable to cross the border, and there is no passenger aviation. Thirdly, alternative transport cannot compensate for the demand. Long-distance buses have lower capacity, less convenience, and slower speeds. Private cars are limited by fuel cost. Finally, UZ has a short booking horizon of only 21 days. It is too short compared to European railways, where tickets are released 30-90 days in advance. As a result, the railway has become the primary mode of transportation with obvious over-demand.

I propose several recommendations based on data exploration and regression models' results. First, it is crucial to increase prices for high-load routes. This segment shows high occupancy and stable sales. They sell out quickly while prices stay low. Increasing fares would help to improve financial stability.

A second recommendation is to introduce early-access tickets. A limited number of wagons could be made available in advance at higher prices. On the one hand, it offers passengers with a high willingness to pay early access and better service conditions. On the other hand, it monetizes extra demand without discriminating against socially sensitive groups.

Third, increasing the booking window from 21 days to 45-60 days would align UZ's system with European practices. Additionally, extended horizons improve the accuracy of demand forecasting and enhance capacity planning.

Finally, wagon allocation can be improved without investments. Since booking patterns are highly predictable, supply can be proactively moved from low-demand to high-demand routes. This would increase sales on high-demand routes and better utilize low-occupancy routes.

The future research can extend these findings. The first step is to compare pre-war and wartime booking curves. This would evaluate how much of the existing over-demand is caused by the war. Secondly, progress requires access to more detailed UZ data, preferably at the individual level. This would allow to use methods from the revenue management literature, such as discrete choice models. Thirdly, it is helpful to check the seasonality and other operational factors. These drivers may additionally explain variations in demand. Finally, future models could combine pricing strategies with wagon allocation. The joint perspective would reflect both constraints: high demand and limited supply. Similar approaches have been proposed in recent studies (e.g., Manchiraju et al., 2025).

REFERENCES

Business Traveller. (2020). Dynamic fares arrive on the Beijing-Shanghai high-speed rail. https://www.businesstraveller.com/business-travel/2020/12/01/dynamic-fares-arrive-on-beijing-shanghai-high-speed-rail/

Cabinet of Ministers of Ukraine. (2025). Resolution No. 1162 of October 23, 2025, "On allocation of funds from the state budget reserve fund." https://zakon.rada.gov.ua/laws/show/1162-2025-p

Cao, Y., Xu, S., Zhu, S., & Wang, Z. (2024). Differentiated passenger demand and supply in high-speed rail networks: A pricing model with route and train-type specificity. PLOS ONE, 19(4), e0314713. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0314713

European Bank for Reconstruction and Development (EBRD). (2023). Ukraine Railways (UZ) Emergency Support: Report of the Supervisory Board. https://www.ebrd.com/content/dam/ebrd_dxp/documents/project/54150/ukraine-railways-uz-emergency-support-board-report.pdf

European Business Association. (2025). Business calls to support Ukrzaliznytsia and include subsidies for passenger transportation in the state budget for 2026. EBA. https://eba.com.ua/en/biznes-zaklykaye-pidtrymaty-at-ukrzaliznytsya-ta-vklyuchyty-vydatky-na-pasazhyrski-perevezennya-do-derzhavnogo-byudzhetu-na-2026-rik/

European Commission. (2023). Integration of the Ukrainian railway system into the EU transport system. Directorate-General for Mobility and Transport. https://transport.ec.europa.eu/system/files/2023-07/Integration_of_the_UAMD_railway_system_into_the_EU_transport_system.pdf

GMK Center. (2024). Ukrainian Railways to raise freight tariffs by 37%: how it will affect the industry. GMK Center. https://gmk.center/en/news/ukrainian-railways-to-raise-freight-tariffs-by-37-how-it-will-affect-the-industry/

Hetrakul, P., & Cirillo, C. (2011). Discrete choice model for Amtrak Acela Express revenue management. https://www.researchgate.net/publication/263327793_Discrete_choice_model_for_Amtrak_Acela_Express_revenue_management

Hetrakul, P., & Cirillo, C. (2013). Accommodating taste heterogeneity in railway passenger choice models based on internet booking data. ResearchGate. https://www.researchgate.net/publication/257694114_Accommodating_taste_heterogeneity_in_railway_passenger_choice_models_based_on_internet_booking_data

Hetrakul, P., & Cirillo, C. (2014). A latent class choice-based model system for railway optimal pricing and seat allocation. ResearchGate. https://www.researchgate.net/publication/259158976_A_latent_class_choice_based_model_system_for_railway_optimal_pricing_and_seat_allocation

Interfax-Ukraine. (2025). Ukrzaliznytsia in 2024 posts UAH 20.4 bln of operating profit from freight transportation, UAH 18.1 bln of loss from passenger transportation. https://en.interfax.com.ua/news/economic/1073694.html

Kyodo News. (2024). Kyushu bullet trains become the first in Japan to adopt dynamic pricing. https://english.kyodonews.net/news/2024/05/5eb2ff6dd685-kyushu-bullet-trains-become-first-in-japan-to-adopt-dynamic-pricing.html

Lethbridge, J. (2019). The Politics of State-Owned Enterprises: The case of the rail sector. https://gala.gre.ac.uk/id/eprint/23207/7/23207%20LETHBRIDGE_The_Politics_of%20State-Owned_Enterprises_2019.pdf

Manchiraju, S., Dawande, M., Janakiraman, G., & Raghunathan, A. (2025). Demand-Driven Pricing and Capacity Planning for Railways. Mitsubishi Electric Research Laboratories (MERL) Technical Report TR2025-007. https://www.merl.com/publications/docs/TR2025-007.pdf

Ministry for Communities, Territories and Infrastructure Development of Ukraine. (2024). Statistical data on Ukrzaliznytsia for 2023–2024. https://mindev.gov.ua/diialnist/napriamy/zaliznychnyi-transport/statystychni-dani-pro-ukrainsku-zaliznytsiu/statystychni-dani-po-at-ukrzaliznytsia-za-2023-2024-roky

Ministry of Infrastructure of Ukraine. (2018). Order No. 184 of April 20, 2018, "On approval of Tariffs for passenger, baggage, and freight transportation by rail within Ukraine." https://zakon.rada.gov.ua/laws/show/z0577-18#Text

Opendatabot. (n.d.). Company profile: Ukrzaliznytsia (JSC "Ukrzaliznytsia"). Contains official registration data, ownership structure, financial reports, and key company statistics. https://opendatabot.ua/c/40075815

PhocusWire. (2024). The smart AI behind effective airline revenue management. https://www.phocuswire.com/pros-smart-science-effective-airline-revenue-management

Price of Travel. (2022). Should you buy European train tickets in advance? Yes, and here's how to save. https://www.priceoftravel.com/buy-european-train-tickets-in-advance

Rail Mitra. (2019). Dynamic fare pricing in premium trains of the Indian Railways. https://medium.com/railmitraaindia/dynamic-fare-pricing-in-premium-trains-of-the-indian-railways-badcf77761cd

State Statistics Service of Ukraine. (n.d.). Statistical information. https://stat.gov.ua/uk/explorer

The Urbanist. (2019). Going the distance on Amtrak. https://www.theurbanist.org/2019/01/16/going-the-distance-on-amtrak/

Vulcano, G., van Ryzin, G., & Chaar, W. (2010). Choice-Based Revenue Management: An Empirical Study of Estimation and Optimization. https://business.columbia.edu/sites/default/files-efs/pubfiles/5396/vulcano%20van%20ryzin%20chaar.pdf

World Bank. (2025a). Restoring and transforming Ukrainian railways for a better future. Feature Story, June 2025. https://www.worldbank.org/en/news/feature/2025/01/16/restoring-and-transforming-ukrainian-railways-for-a-better-future

World Bank. (2025b). Ukraine's Transport and Logistics System: Current and Prospective Opportunities and Challenges. January 16, 2025. https://documents1.worldbank.org/curated/en/099061725033525342/pdf/P502442-346a4fd3-882f-46ca-95c9-ce90c0a71619.pdf

APPENDIX A. TABLES

Table 1 – Descriptive statistics of key variables

variable	count	mean	std	min	25%	median	75%	max
price	440853	508.89	256.72	81	314	482	717	1251
seat_count	440853	53.00	18.22	20	40	54	56	134
available_seats	440853	26.60	22.75	1	8	22	39	131
seats_booked	440853	26.40	16.11	0	13	26	35	133
price_per_km	440853	0.92	0.35	0.24	0.59	0.91	1.16	2

Table 2 – Descriptive statistics by route

route	count	mean price	median price	mean occupancy	median occupancy
Zaporizhzhia → Kyiv	45221	541.58	502	0.53	0.55
Kyiv → Zaporizhzhia	39053	551.57	517	0.49	0.47
Kharkiv → Kyiv	35419	514.96	489	0.49	0.48
Kyiv → Kharkiv	26091	544.86	489	0.42	0.38
Kyiv → Dnipro	21736	595.04	517	0.35	0.28
Lviv → Kharkiv	17307	626.62	613	0.80	0.85
Dnipro → Kyiv	16760	600.77	517	0.35	0.30
Kyiv → Sumy	14216	170.64	171	0.43	0.39
Kharkiv → Lviv	12948	579.73	387	0.77	0.82
Lviv → Zaporizhzhia	12051	232.57	225	0.54	0.50
Lviv → Dnipro	10415	878.21	982	0.59	0.68
Odesa → Kyiv	10414	690.69	791	0.57	0.60

route	count	mean price	median price	mean occupancy	median occupancy
Zaporizhzhia → Lviv	9666	211.72	220	0.50	0.46
Sumy → Kyiv	9567	172.05	171	0.54	0.53
Lviv → Odesa	9382	697.62	791	0.55	0.60
Lviv → Darnytsia	8627	599.05	498	0.43	0.39
Kyiv → Odesa	8026	714.84	791	0.58	0.62
Mukachevo → Odesa	7653	313.73	337	0.73	0.78
Odesa → Lviv	7516	712.48	791	0.59	0.65
Darnytsia → Lviv	7031	603.38	545	0.49	0.45

Table 3 – Descriptive statistics by class

route	count	mean price	median price	mean occupancy	median occupancy
Regular coupe	127943	551.77	412	0.70	0.78
2nd class	124119	450.67	472	0.48	0.46
1st class	112596	662.29	727	0.37	0.30
Platzkart	47473	231.68	209	0.63	0.69
3rd class	13500	95.76	95	0.12	0.09
Women coupe	8021	712.73	681	0.84	0.88
Child coupe	7200	726.85	678.00	0.84	0.89

Table 4 – H1 models comparison

Specification	β_1	β_2	**Pivot ($-\beta_1/2\beta_2$)**	**R^2 (within)**	**F-statistic**
Log-Log	0.0978	–	–	0.002	103.1
Level-Quadratic	-5.1021	0.3104	8.22	0.057	1321.9

Specification	β_1	β_2	Pivot ($-\beta_1/2\beta_2$)	R^2 (within)	F-statistic
Log-Quadratic	-0.1897	0.0112	8.51	0.073	1714.8

Notes: All specifications estimated with entity fixed effects and clustered SEs by entity. Pivot reported for quadratic specifications only. * p<0.10, ** p<0.05, *** p<0.01.

Table 5 – Determinants of final occupancy (Pooled OLS)

Constant	3.244*** (0.341)
Weekend	0.023 (0.061)
Holiday	-0.109 (0.101)
Journey hours	-0.093*** (0.020)
R2 (adj)	0.394
F-statistic	61.21

Notes: Pooled OLS with HC3 robust SE in parentheses; * p<0.10, ** p<0.05, *** p<0.01. The dependent variable is the logit-transformed final occupancy rate.

Table 6 – Volatility of daily sales vs. final load

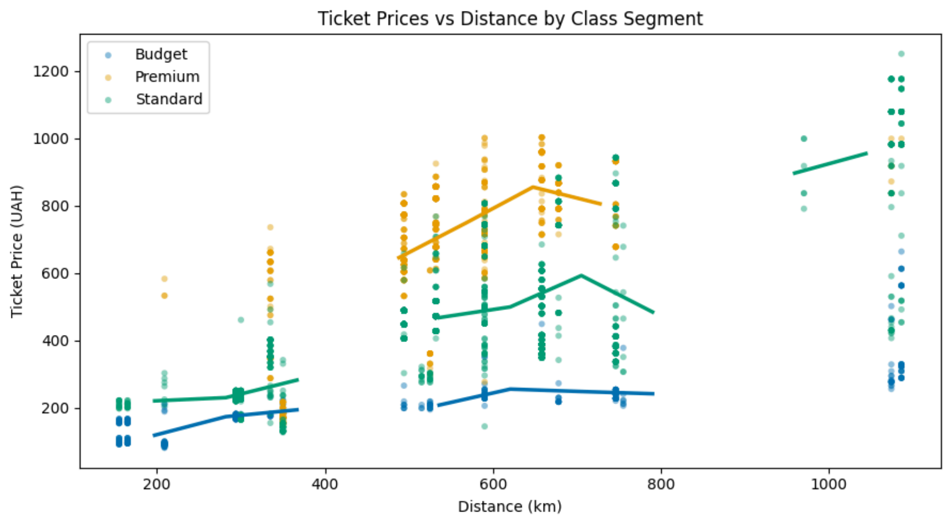
Metric	Value
Mean variance (high-load)	3024
Mean variance (low-load)	2009
Ratio (high/low)	1.5
t-test (Welch) – t, p	3.491, 0.0005
Mann-Whitney – U, p	573854, 0.0000

Table 7 – Stability of booking curves

Metric	**Value**
Tau (threshold)	0.2
Segments	144
Median width	0.006
Share below tau	0.924
Wilcoxon (stat, p, n)	281, 0.0000, 144

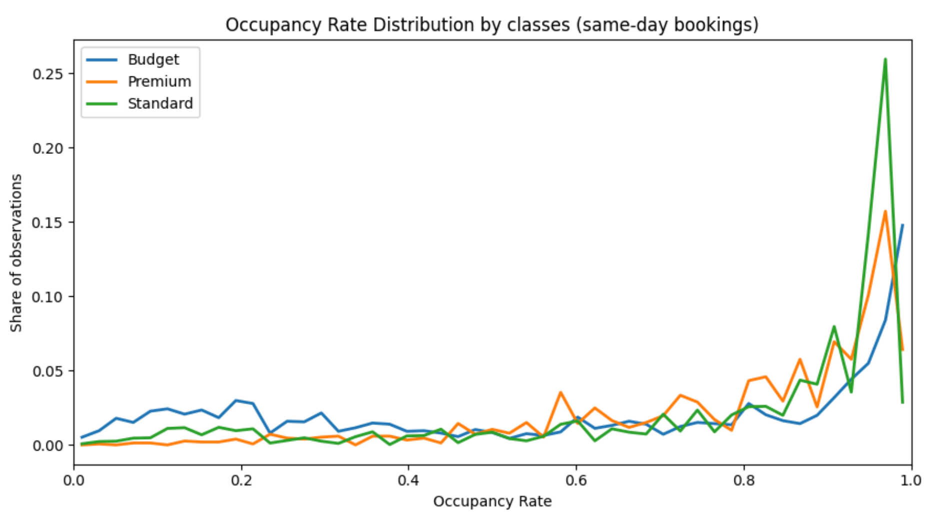
APPENDIX B. FIGURES

Figure 1 – Prices and distance by class segment



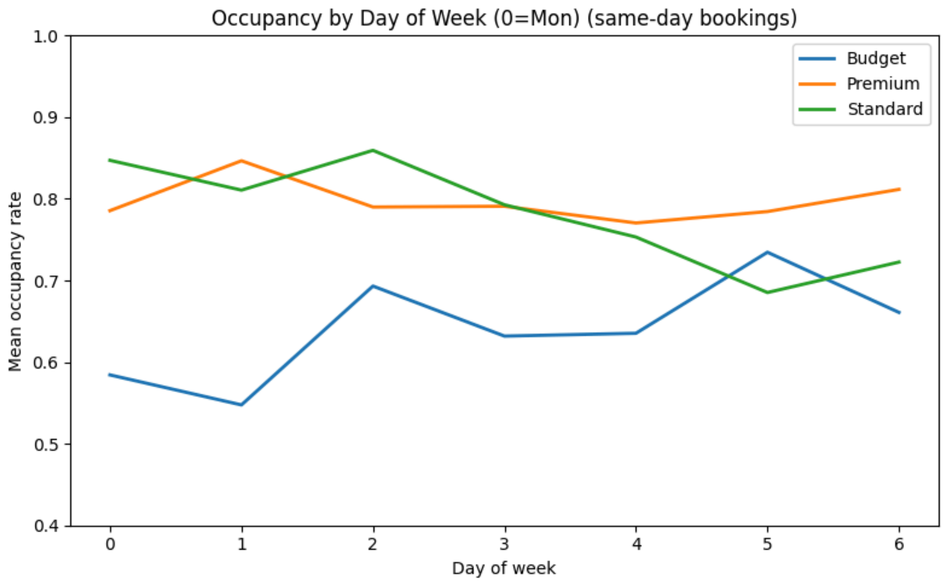
Note: Budget (Platzkart, 3rd class), Standard (Coupe, 2nd class), Premium (Luxury, 1st class, Female/Child coupe)
Source: Author's calculations based on UZ scraped data

Figure 2 – Final occupancy rate distribution by class segment



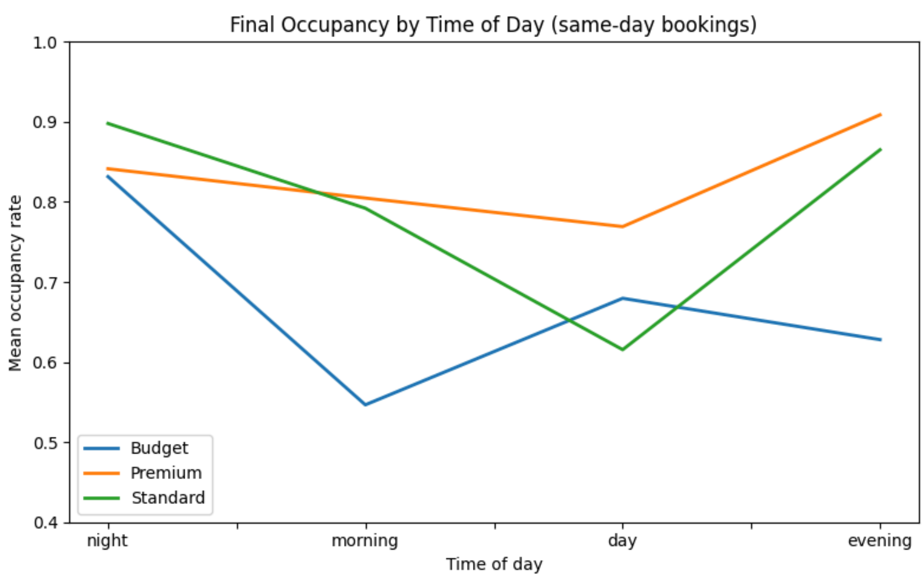
Note: Budget (Platzkart, 3rd class), Standard (Coupe, 2nd class), Premium (Luxury, 1st class, Female/Child coupe)
Source: Author's calculations based on UZ scraped data

Figure 3 – Final occupancy by day of the week and class segment



Note: Budget (Platzkart, 3rd class), Standard (Coupe, 2nd class), Premium (Luxury, 1st class, Female/Child coupe)
Source: Author's calculations based on UZ scraped data

Figure 4 – Final occupancy by time of the day and class segment



Note: Budget (Platzkart, 3rd class), Standard (Coupe, 2nd class), Premium (Luxury, 1st class, Female/Child coupe)
Source: Author's calculations based on UZ scraped data

Figure 5 – Cumulative booking curves by class

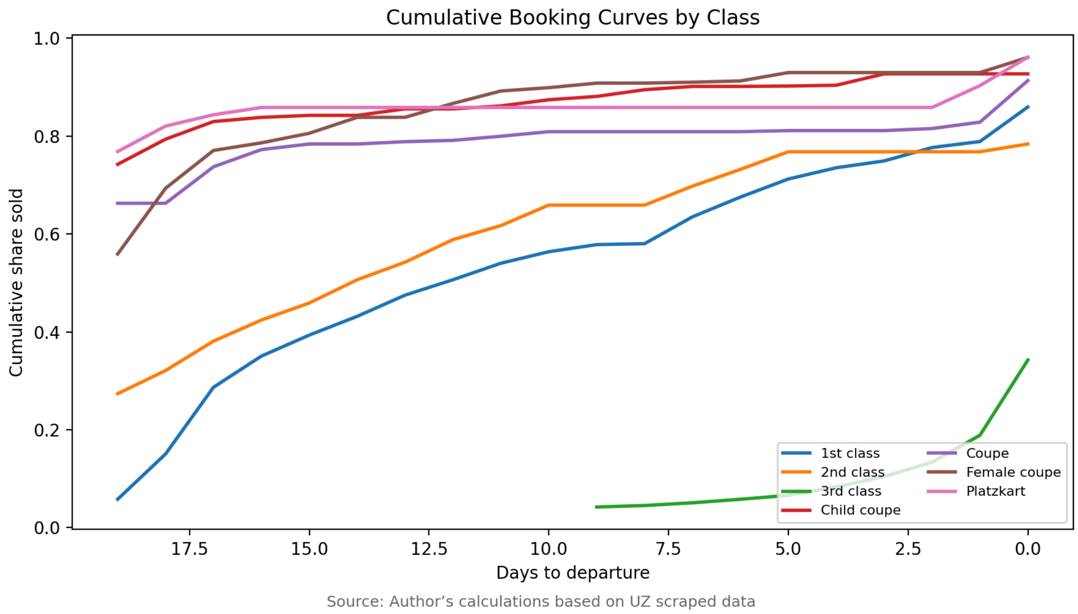


Figure 6 – Daily booking curves by class

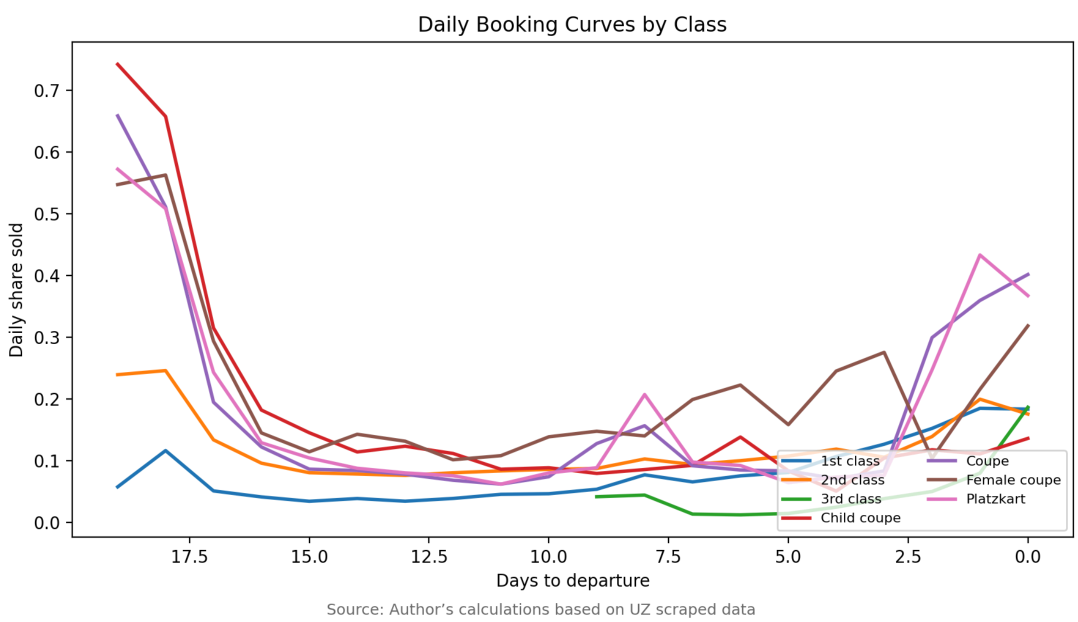


Figure 7 – Estimated U-shaped booking curve (Log-Quadratic FE)

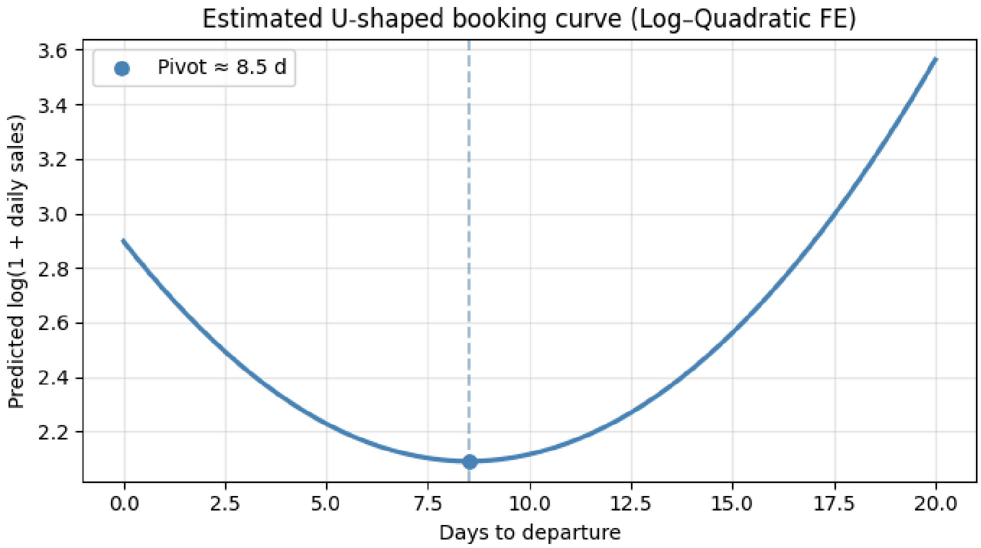


Figure 8 – Estimated U-shaped booking curve (Level-Quadratic FE)