Market Interaction between Rail and Air:

The Role of Train Pricing in U.S. Domestic Flight Demand

Stina Andersson

Degree project/Independent project • 15 credits
Swedish University of Agricultural Sciences, SLU
Faculty of Natural Resources and Agricultural Sciences/Department of Economics
Business and Economics – Sustainable Development
Degree project/SLU, Department of Economics, 1683 • ISSN 1401-4084
Uppsala 2025

Market Interaction Between Rail and Air: The Role of Train Pricing in U.S. Domestic Flight Demand

Stina Andersson

Supervisor:	**Levi Soborowicz, Swedish University of Agricultural Science, Department of Economics**
Examiner:	Shon Ferguson, Swedish University of Agricultural Science, Department of Economics

Credits:	15 credits
Level:	First cycle, G2E
Course title:	Independent project in Economics
Course code:	EX0903
Programme/education:	Business and Economics – Sustainable Development
Course coordinating dept:	Department of Economics
Place of publication:	Uppsala
Year of publication:	2025
Copyright:	All featured images are used with permission from the copyright owner.
Title of series:	Degree project/SLU, Department of Economics
Part number:	1683
ISSN:	1401-4084

Keywords:	Cross-price elasticity, Demand estimation, Modal shift, First difference estimator, Air travel demand, Intercity transport, Train fares.

Swedish University of Agricultural Sciences
Faculty of Natural Resources and Agricultural Sciences
Department of Economics

Abstract

The United States's transport sector accounts for 28 % of national CO_2 emissions. Since U.S. intercity trains emit approximately 39% less CO_2 per passenger mile than domestic flights, understanding the potential for modal shift from air to rail is crucial for climate policy. This thesis investigates how train ticket prices influences the demand for domestic flights in the U.S., i.e. the cross-price elasticity between the two modes of transportation. The analysis uses annual panel data covering six domestic routes over the time period 2013-2023. A log-log demand model for air travel is estimated using a first difference estimator with a linear constraint, helping to control for time-invariant route-specific factors and address endogeneity. The results indicate a small, positive, but statistically insignificant, cross-price elasticity, suggesting limited substitution from air to rail in the U.S. market. In contrast, the own-price elasticity of flight demand is statistically significant and inelastic (approximately -0.94). These findings imply that policy efforts to promote modal shift from air to rail will likely require more than adjustments in rail pricing alone. Instead direct pricing interventions targeting air travel, such as taxes or fees, are likely to be more effective.

Keywords: Cross-price elasticity, Demand estimation, Modal shift, First difference estimator, Air travel demand, Intercity transport, Train fares.

Table of contents

List of tables

1. Introduction

In a world struggling with increasing levels of greenhouse-gas emissions (Ritchie et al. 2024) it is crucial to identify the principal emission sources and determine the most effective measures for their reduction. In the United States, the transport sector alone generates 28 percent of national CO_2 emissions, making it the largest contributor of greenhouse gases in the country (US EPA 2025). The climate impact from transportation differs depending on transportation mode, energy mix, transportation technology and occupancy. For example, the average national train emits 14% of the carbon of a domestic flight (Ritchie 2023).

In the U.S., intercity rail traffic is operated by the federally owned monopoly Amtrak (AAR 2025). Most of these trains are diesel-powered (Amtrak 2024) and emit approximately 0.15 kg of CO_2 per passenger mile (Amtrak 2022), roughly 39 % less than domestic flights (Ritchie 2023). Despite these environmental advantages, trains only play a marginal role in U.S. long-distance travel. According to the Bureau of Transportation (2017), 90% of long-distance trips (defined as those extending at least 50 miles) are made by private car, approximately 7% by airplane, 2% by bus, and only 1% by train. The very low share of rail travel indicates potential to increase train ridership through targeted policy interventions aimed at promoting a modal shift from air to rail. Against this background, this thesis investigates the following research question: How do train ticket prices affect the demand for domestic flights in the U.S.?

IPCC (2023:1060) states that shifting travel-mode demand for urban and intercity transport is crucial for decarbonizing the transport sector. They emphasize that policymakers must understand the price relationship between modes of transportation to design effective modal-shift policies. Yet empirical estimates of the cross-price elasticity between rail and air travel remain scarce. The cross-price elasticity represents the percentage change in the quantity demanded of one good in response to a one percent change in the price of another good.

Gama (2017) is among the few studies that empirically estimate the cross-price elasticity between U.S. train fares and air travel demand. She applies a discrete-choice model (Berry 1994; Berry et al. 1995) to monthly data from October 2009 through September 2010. This period is relevant because it corresponds to President Obama's announcement of a $13 billion passenger-rail investment. To address endogeneity, because prices are correlated with unobserved factors that affect demand, Gama instruments fares with cost shifters such as jet-fuel spot prices for air fares and diesel spot prices for rail fares. She finds that air travel's own-price elasticity is roughly four times that of rail, and that a 1% increase in train fares raises airline market share by just 0.0008%. More recently, Escañuela Romana et al. (2023a) estimate intercity demand elasticities using the Rotterdam Demand Model (RDM) applied to annual U.S. data from 2003 to 2019. They find that the Marshallian demand elasticities for air, road, and rail travel are all inelastic, and that the estimated cross-price elasticities between modes are weak or statistically insignificant. The authors conclude that additional research is needed to better understand the behavioral drivers of rail travel and other explanatory factors, in order to obtain more robust estimates of rail transport elasticities.

Other research methods have been employed to answer similar questions. Zeng et al. (2021) employ a stated-preference survey in China in 2020 and estimate a rail–air cross-price elasticity of 0.000123, and simulate resulting CO_2-savings under various pricing schemes. No consideration is given to the fact that their survey was distributed at the beginning of the COVID-19 pandemic. In a study conducted by Wardman and Tyler (2000) the impact of rail network accessibility on inter-urban demand is examined. The authors conclude that accessibility plays a relatively minor role, with journey distance, prior service experience, and fare levels exerting far greater influence on train ridership.

There are studies specifically addressing the own-price elasticity of long-distance rail demand. For example, Börjesson (2014) estimates the elasticity at -0.72 for business trips and -0.59 for private trips in Sweden and Rohr et al. (2013) estimate that the UK long-distance rail fare elasticities vary between -0.34 and -0.76. Regarding the own-price elasticity of demand for domestic flights on U.S. routes, it's estimated to be inelastic, with a value of -0.70 (Escañuela Romana et al. 2023b). Brons et al. (2002), in a meta-analysis covering multiple countries, further find that long-run own-price elasticities for air travel tend to exceed short-run values. They also conclude that available substitutes (including rail) have a limited impact on price sensitivity. Unfortunately, their analysis omits income as a control variable, which they acknowledge may bias the estimated elasticities. The income elasticity of air travel is instead investigated by Gallet & Doucouliagos (2014) and they estimate the income elasticity of domestic flights at 1.186. The authors find that income elasticity is higher for international flights than for domestic, but is relatively stable across geographical contexts.

Building on the limited U.S. evidence this thesis estimates the cross-price elasticity between rail and air by exploiting annual data from 2013 to 2023 and thus captures both long-run behavioral adjustments and the COVID-19 shock. Existing U.S.-based studies either focus on short timeframes around specific policy events or are limited to the pre-pandemic period, leaving the longer-term dynamics and pandemic-related shifts in travel behavior underexplored. There is also a notable gap in European-level studies on the substitution between rail and air, primarily due to limited availability of relevant data. Since this study focuses on U.S. corridors in the Northeast, Midwest and Southeast, which share key characteristics with major European regions (e.g. population density, travel distances and development of rail infrastructure), it aims to provide insights that are also relevant to Western Europe.

To address endogeneity, I initially considered using diesel and jet-fuel prices as instruments, following Gama (2017). However, both fuel prices are driven by the global crude-oil market, which also directly affects airline operating costs and ticket fares. This relationship violates the exclusion restriction for an instrumental variable. Instead, I address endogeneity through two avenues in this thesis. First, I estimate the parameters using a first differencing estimator, in order to remove route-specific unobservable heterogeneity. Second, a set of relevant controls is included, for example gasoline prices, an annual income control, rail on-time performance metrics to capture service-level effects, and COVID-19 dummies to isolate pandemic disruptions as the major time-varying shock. Estimating the first difference model with a linear constraint reveals a small and positive, statistically insignificant, cross-price elasticity, suggesting limited substitution from air to rail in the U.S. market.

This thesis is organized as follows: The theoretical background is described in Section 2. Data sources, variable constructions and definitions are presented in Section 3. The methodology is explained in Section 4, including the motivation for using a first difference estimator and the treatment of Covid-related shocks. The results are presented in Section 5, followed by a discussion in Section 6 that analysis findings, consider potential limitations and offers suggestions for future research. This ends with a conclusion in Section 7.

2. Theoretical Background

2.1 Utility Maximization and Demand

When it comes to modeling consumer demand, utility maximization provides the main microeconomic foundation (Jevons 1874). In the context of travel, individuals choose combinations of transportation modes to maximize their utility, subject to resource constraints. When applied to travel decisions, this optimization problem yields demand functions that relate trip quantities to the price of substitutes and complements, traveler's income and preferences, as well as service attributes. A utility maximization problem for two goods can be expressed as:

$$Max_{x,y}\ U(x,y)\ \ st.\ \ I\ =\ P_x x +\ P_y y$$

where the utility U is a function of x and y (Nicholson & Snyder 2010:123).

According to the first order condition (FOC) of the utility maximization problem, a consumer chooses a combination of goods such that the marginal rate of substitution (MRS) between the two goods equals their relative price ratio (Autor 2016). This condition holds for interior solutions, where the consumer purchases strictly positive amounts of both goods. At this point, under the assumption of nonsatiation, the consumer cannot increase utility further by reallocating expenditure between the goods.

$$First\ order\ condition: \qquad \frac{MU_x}{MU_y} = \frac{P_x}{P_y} =\ MRS\ (x\ for\ y)$$

Solving the utility maximization problem yields Marshallian demand functions, where the quantity demanded of a good depends on its own price, the price of the other good and the consumer's income (Nicholson & Snyder 2010:145).

$$Marshallian\ demand: \qquad x_i = f\left(P_x, P_y, I\right)$$

2.2 Substitution and Income Effect

A price change results in two effects: a substitution effect and an income effect (Nicholson & Snyder 2010:149-150). When the price of a good changes, the relative prices of goods are altered, which influences the rate at which one good can be substituted for another; i.e. the MRS.

The substitution effect captures how a consumer reallocates consumption to relatively cheaper goods when the price change occurs, holding utility constant. Thereby it reflects the pure price effect, isolated from any change in real income. This effect is always negative; as the price of a good increases, the consumer substitutes away from it (Varian 2010:142).

In addition, the change in price alters a consumer's real income, generating an income effect. This occurs because a price increase reduces the consumer's ability to purchase the same bundle as before, whereas a price decrease increases it (Nicholson & Snyder, 2010:149-150). As a result, the consumer moves to a new indifference curve. The income effect captures the portion of the total change in quantity demanded that is attributable to the change in real income. The direction and magnitude of this effect depend on the nature of the good (whether it is normal, inferior, or Giffen).

2.3 Slutsky Equation

The Slutsky equation (Slutsky 1915) provides the theoretical framework for decomposing the total effect of a price change into a substitution effect and income effect, as described in the previous section.

$$Slutsky: \quad \frac{\partial x\left(p_x, p_y, I\right)}{\partial p_x} = \left.\frac{\partial x}{\partial p_x}\right|_{U=const.} - x\frac{\partial x}{\partial I}$$
$$= substitution\ effect + income\ effect$$

Although originally developed for analyzing own-price effects, this reasoning also helps explain how shifts in relative prices between substitute goods may influence demand. In the context of train prices and flight demand, a decrease in train fares may induce consumers to substitution away from air travel, as rail becomes relatively cheaper. At the same time, the lower price increases real income, potentially affecting demand via the income effect. To isolate this income-related mechanism, GDP per capita is included as a control variable in the model. If air travel is a normal good, the income effect is expected to be positive (i.e. $\frac{\partial X}{\partial I} > 0$).

2.4 Elasticities

In the context of this study, elasticities are unitless measures used to capture how the quantity demanded responds to changes in price and income. Specifically, I use elasticities to quantify the impact of train ticket prices on the demand for air travel. Elasticities can be calculated through a logarithmic transformation of the standard linear regression model. That is, when both the dependent and independent variables are transformed using the natural logarithm function. Then the estimated coefficient directly represents the percentage change in the dependent variable associated with a one-percent change in the independent variable, i.e. the elasticity. Since price variables are expressed in nominal terms and income is measured in real terms (chained GDP per capita in 2017 dollars), the estimated elasticities reflect Marshallian (uncompensated) demand. That is, demand responses are not compensated to hold utility constant, and thus capture both substitution and income effects (Nicholson & Snyder, 2010:158). Below is a brief overview of the three elasticities estimated in this thesis.

2.4.1 Own-Price Elasticity of Demand

The own-price elasticity of demand measures the percentage change in quantity demanded in response to the percentage change in a good's own price, ceteris paribus (Nicholson & Snyder, 2010:162-163). Thus, it captures both substitution and income effects. A negative own-price elasticity (e_P< 0) indicates that the good is normal, meaning that demand increases as the price decreases. Accounting for the own-price elasticity of demand is essential in this study, as it provides insight into how flight demand responds to changes in air travel fare.

$$e_P = \frac{\partial x(p_x, p_y, I)}{\partial p_x} * \frac{p_x}{x}$$

2.4.2 Income Elasticity of Demand

The income elasticity of demand measures the percentage change in quantity demanded in response to a percentage change in the consumer's income, ceteris paribus (Nicholson & Snyder 2010:162-163). A positive income elasticity ($e_I > 0$) indicates that the good is normal and that demand increases as the income increases.

$$e_I = \frac{\partial x(p_x, p_y, I)}{\partial I} * \frac{I}{x}$$

2.4.3 Cross-Price Elasticity

The cross-price elasticity of demand measures the percentage change in the quantity demanded of good x in response to a percentage change in the price of y, ceteris paribus (Nicholson & Snyder, 2010:190). A positive cross-price elasticity ($e_{x,p_y} > 0$) indicates that the goods are substitutes, since an increase in the price of one good increase the demand for the other good. On the other hand, if the cross-price elasticity is negative, the goods are complements and if it's zero the goods are independent of each other. The cross-price elasticity is the main focus in this thesis since it captures how changes in the train price affect the demand for flights.

$$e_{x,y} = \frac{\partial x(p_x, p_y, I)}{\partial p_y} * \frac{p_y}{x}$$

2.5 Hypotheses

Based on the theoretical background there are five testable hypotheses:

H1: $e_{f,t} > 0$. The cross-price elasticity of demand between flights and train ticket prices is expected to be positive, as long-distance rail travel is typically considered a substitute for domestic air travel.

H2: $e_P < 0$. The own-price elasticity of demand for flights is expected to be negative, as higher airfares are likely to reduce the quantity of air travel demanded.

H3: $e_{f,g} > 0$. Traveling by car is a common substitute for domestic air travel. Therefore, an increase in gasoline prices is expected to raise the demand for flights, implying a positive cross-price elasticity between flight demand and gasoline prices.

H4: $e_I > 0$. Higher income increases consumers' purchasing power and typically leads to greater demand for discretionary goods, such as air travel. Therefore, the income elasticity of demand for flights is expected to be positive. In this study, GDP per capita is used as a proxy for income.

H5: $\beta_1 < 0$. Higher on-time performance increases perceived quality of rail service, which is expected to raise train demand and to reduce demand for domestic air travel. This implies a negative relationship between Amtrak’s on-time performance and air travel demand if rail and air are substitutes.

3. Data & Variable Construction

The data consist of an annual panel covering the period from 2013 to 2024. Six domestic routes from the Northeast, Midwest, and Southeast regions, where both train and airline services operate and offer comparable services, are included. The routes are: Boston–Cleveland, Cleveland–Chicago, Chicago–Indianapolis, Philadelphia–Washington, Richmond–Charlotte, and Washington–Cincinnati. Each route is defined as travel in either direction between the two cities. This yields a total of 72 observations, with a balanced panel from 2013 to 2023. Since on-time performance (OTP) data for 2024 are missing, that year is excluded in specifications, including the OTP variable, resulting in a reduced sample of 66 observations. All variables are measured or converted to annual frequency.

The dependent variable is defined as the annual number of passengers per route. It is constructed by aggregating quarterly domestic flight passenger data from the U.S. Department of Transportation (2025).

Several variables reflect travel-related prices for elasticity estimation. Since Amtrak does not publish fare data, I approximate train ticket prices using available national-level revenue data. Specifically, I use Amtrak's annual passenger-related revenue per passenger mile as a proxy for the average fare per mile. This value is then multiplied by the distance of each route i to estimate the total fare:

$$\mathbf{P}^{\text{train}} = \frac{Annual\ Passenger\ Revenue_t}{Total\ Passenger\ Miles_t} * Distance_i$$

The approximation is motivated by microeconomic theory:

$$AR = \frac{TR}{Q} = \frac{P * Q}{Q} = P$$

which shows that average revenue (AR) equals the market price (P), under the assumption of a single price per unit. In this context, the average revenue per passenger mile is assumed to represent the average fare per mile paid by passengers. This also implies symmetric pricing, i.e., that fares are the same in both direcations along each route. While this approach provides a consistent estimate of fares across time and routes, it has several limitations. First, the revenue per passenger mile is a national-level average and does not reflect route-specific pricing, which for example can vary due to local demand, competition, and service characteristics. Second, Amtrak applies dynamic pricing, meaning that fares tend to differ depending on booking time (Amtrak 2025b) and variations like that are not captured by the average revenue. Third, multiplying by route distance assumes a linear fare structure, although actual pricing may involve fixed components or discounts.

Despite these limitations, the proxy remains the most feasible option given the lack of disaggregated fare data and allows for comparability across routes and years. While actual transaction-level fare data in some respects might be preferable, it could also introduce other types of endogeneity, such as those related to individual booking behavior and unobserved demand factors. By relying on an aggregated pricing structure, the proxy may mitigate some of these concerns, though it introduces its own trade-offs.

The data used to construct the train fare proxy comes from three sources: revenue data are sourced from Amtrak's annual reports (Amtrak 2025a), passenger miles from the Federal Railroad Administration (n.d.), and route distances from fact sheets provided by the Rail Passengers Association (n.d.).

Flight fares are calculated as volume-weighted annual averages using quarterly fare and passenger volume data (U.S. Department of Transportation 2025). Gasoline prices represent the national average retail price across all grades and formulations, obtained from the U.S. Energy Information Administration (EIA 2025).

Additional control variables account for macroeconomic and service-related factors. Real GDP per capita, expressed in chained 2017 dollars and seasonally adjusted at an annual rate, is sourced from the Federal Reserve Bank of St. Louis (FRED 2025). Amtrak's on-time performance (OTP) is defined as the percentage of trains arriving within 15 minutes of the scheduled time (Amtrak n.d.). This variable covers the years 2013–2023 and is weighted by distance category (Bureau of Transportation Statistics 2024). Finally, a COVID-19 dummy variable takes the value 1 for the years 2020 to 2022, and 0 otherwise, based on the World Health Organization's classification of the pandemic period as a public health emergency of international concern (World Health Organization 2023).

Table 1 below summarizes the variables used in the study, including definitions and data sources. Table 2 provides summary statistics of the data. A correlation matrix over the data can be found in Appendix A.

Table 1. Variable definitions and data sources

Variables	**Definition**	**Source**
$\mathbf{Q}^{\text{flight}}$	Annual number of passengers per Route.	U.S. Department of Transportation (2025)
$\mathbf{P}^{\text{train}}$	Proxy for train ticket prices. Total passenger related revenue (millions of dollars) / Passenger miles (millions of miles) * Route distance (miles).	Amtrak (2025a) Federal Railroad Administration (n.d.) Rail Passengers Association (n.d.)
$\mathbf{P}^{\text{flight}}$	Volume weighted annual average flight fare per route (dollars).	U.S. Department of Transportation (2025).
$\mathbf{P}^{\text{gas}}$	U.S. All Grades All Formulations Retail Gasoline Prices (Dollars per Gallon).	U.S. Energy Information Administration (EIA 2025)
GDP	Real gross domestic product per capita. Chained 2017 Dollars, Seasonally Adjusted Annual Rate.	Federal Reserve Bank of St. Louis (FRED 2025)
OTP	Amtrak's On-Time Performance (percentage). Weighted by distance.	U.S. Bureau of Transportation Statistics (2024)
I	Dummy variable representing the Covid-19 pandemic. It takes the value 1 for years 2020–2022, and 0 otherwise.	World Health Organization (2023)

Table 2. Summary statistics

	Mean	SD	Min	Max	Total observations	Number of Years	Number of routes
Q flight	160,722	145,689	1,547	487,377	72	12	6
P train	$148.22	$81.86	$33.37	$384.53	72	12	6
P flight	$216.72	$52.50	$142.11	$363.75	72	12	6
P gas	$3.02	$.58	$2.25	$4.06	72	12	6
GDP	$61,842	$3,871	$56,172	$68,501	72	12	6
OTP	75.78	3.39	71.20	82.3	66	11	6

4. Methodology

4.1 Econometric Specification

4.1.1 Levels Model

The model stated esimates how train ticket prices influence the demand for domestics flights in the U.S., while controlling for other factors that may affect passenger volumes.

$$ln(Q^{\text{flight}}{}_{it}) = \beta_0 + e_{f,t}\, ln(P^{\text{train}}{}_{it}\,) + e_P\, ln(P^{\text{flight}}{}_{it}) + e_{f,g}\, ln(P^{\text{gas}}{}_{t}) \\ + e_I\, ln(GDP_t) + \beta_1 OTP_t + C \cdot I(Covid\ =\ 1) + \alpha_i + u_{it}$$

4.1.2 Variable Rationale

The continuous variables are transformed to logarithms to enable interpretations in terms of elasticities, as described in Section 2.4, Elasticities. The variables are indexed by route *i* and/or year *t*.

The dependent variable, $ln(Q^{\text{flight}}{}_{it})$, represents the annual number of airline passengers on a given route and year, serving as a measure of air travel demand. The main variable of interest is $ln(P^{\text{train}}{}_{it})$, which serves as a proxy for train ticket prices (see Section 3., Data & Variable Construction). This variable is included in the model to estimate the cross-price elasticity of flight demand with respect to train fares. The corresponding coefficient, $e_{f,t}$, quantifies the directional and proportional effect that changes in rail ticket prices have on air travel demand. A positive and statistically significant estimate would indicate that rail and air travel are substitutes.

To isolate this effect, the model includes $ln(P^{\text{flight}}{}_{it})$, the own-price of air travel. The corresponding coefficient e_P captures the own-price elasticity of flight demand. A negative and statistically significant estimate would be consistent with a normal good, indicating that higher flight prices reduce demand for air travel. In addition, $ln(P^{\text{gas}}{}_{t})$, the national average gasoline price, is included to control for the cost of car travel, which is another substitute for both air and train travel. Given that 92 percent of vehicles in the U.S. run on gasoline (IER 2023), this variable serves as a reasonable proxy for the cost of driving. A positive value of the coefficient $e_{f,g}$ would suggest a positive cross-price elasticity, and that higher gasoline prices increase the demand for alternative transport modes such as air travel.

Next, $ln(GDP_t)$, representing the real gross domestic product per capita, is included to control for purchasing power and broader macroeconomic conditions that may affect travel demand independently of transport prices. As income levels rise, demand for air travel may increase regardless of relative costs, making this a crucial control variable. Service quality is captured by Amtrak's on-time performance, OTP_t. Punctuality is highly valued by travelers (U.S. Department of Transportation n.d.) and variation in OTP may influence the relative attractiveness of rail versus air, even when prices are held constant.

To account for the large time-varying shock the COVID-19 pandemic resulted in during the sample period, the model includes the dummy variable *I*. The dummy equals 1 if the observation falls within the years 2020 to 2022, and 0 otherwise. In doing so, the associated coefficient, C, captures the average change in air travel demand during the pandemic period relative to non-pandemic years, holding other covariates constant. Lastly, the variable α_i denotes time-invariant route fixed effects, accounting for unobserved heterogeneity across routes such as geography, underlying service levels, or consistent patterns in traveler demand. u_{it} is the idiosyncratic error term, capturing all remaining variation in $ln(Q^{\text{flight}}{}_{it})$ over time and across routes.

4.2 Identification Strategy

4.2.1 First Differences Estimator

The first differences estimator is defined by (Wooldridge 2002:279):

$$\Delta y_{it} = \Delta x_{it}\beta + \Delta u_{it}$$

where: $\Delta y_{it} = y_{it} - y_{i,t-1}$, $\Delta x_{it} = x_{it} - x_{i,t-1}$ & $\Delta u_{it} = u_{it} - u_{i,t-1} = e_{it}$

In this study, a log-log air travel demand model is estimated using a first differences estimator with a linear constraint:

$$\begin{aligned}\Delta ln(Q^{\text{flight}}{}_{it}) = {} & e_{f,t}\,\Delta ln(P^{\text{train}}{}_{it}) + e_P\,\Delta ln(P^{\text{flight}}{}_{it}) + e_{f,g}\,\Delta ln(P^{\text{gas}}{}_{t}) \\ & + e_I\,\Delta ln(GDP_t) + \beta_1\,\Delta OTP_t + \beta_2 CovidB + \beta_3 CovidE \\ & + \Delta u_{it}\ s.t\ \ \beta_2 + \beta_3 = 0\end{aligned}$$

The model is estimated in Stata 18 using the built-in D. operator to generate the first differences of each variable. The first period for each cross section is lost due to the differencing, leaving $t - 1$ time periods for each route *i*. This results in 60 total observations left over for when the OTP variable is included. The estimation is run with clustered standard errors at the route level, to account for arbitrary heteroskedasticity and serial correlation within each route.

4.2.2 Motivation and Intuition Behind First Differences

In an ideal setting, train fares would be exogenous, i.e. uncorrelated, with both flight demand and any other determinants of flight demand, so that $Cov(P^{\text{train}}{}_{it}, u_{it}) = 0$. Then a simple OLS would be unbiased. In reality, however, passengers compare rail and air (e.g. via booking sites), making fare levels, and especially fare changes, endogenous to factors driving demand for both modes. Moreover, both rail and air transportation are often subject to common shocks such as changes in global fuel prices, which simultaneously affect operating costs and thus ticket prices across modes. This simultaneity introduces a correlation between price variables and the error term, further violating the exogeneity condition. To eliminate these confounding factors, I take the first differences of the variables, by modeling each series as a random walk. This eliminates all time-invariant factors i.e. route unobservable effects (α_i) and the change in error term, Δu_{it}, between time t and $t-1$, becomes uncorrelated with the covariates (Wooldridge 2002:279–281). The random walk assumption is denoted:

$$u_{it} = u_{i,t-1} + e_{it} \rightarrow \Delta u_{it} = e_{it}$$

where e_{it} is a random shock term that is independently and identically distributed with mean zero and variance σ_e^2 (i.e. $e_{it} \sim$ i.i.d. $(0, \sigma_e^2)$) (Wooldridge 2013:391). This random walk causes Δu_{it} to be serially uncorrelated with the differenced covariates (Wooldridge 2013:482):

$$E[(u_{it} - u_{i,t-1})(x_{it} - x_{i,t-1})] = 0 \text{ meaning that } E(\Delta u_{it} \mid X_i) = 0$$

Moreover, a random walk satisfies $E(x_{t+h}|x_t) = x_t$ for all h ≥ 1, when h denotes periods hence, which means that the value of x today is the best predictor of tomorrow's (or any future) value of x (Wooldridge 2013:392). From that it follows that $\Delta x_{it} = x_{it} - x_{i,t-1}$ captures unpredictable innovation, which makes Δx_{it} behave like a "as if randomized" variable. This satisfies the identifying assumption that enables a causal identification of the effect from train prices on flight demand.

An alternative to the first difference estimator would be a fixed effects (FE) estimator. However, in this context, the Wooldridge test for serial correlation is highly significant (p-value < 0.01), which suggests that the standard assumptions required for the FE estimator may not hold (Wooldridge 2002; Drukker 2003). Specifically, the efficiency of the FE estimator relies on the assumption of homoskedasticity and no serial correlation in the idiosyncratic errors, i.e. $E(u_i u_i' \mid x_i, c_i) = \sigma_u^2 I_T$ (Wooldridge 2002:269). When this assumption is violated, as indicated by the test result, the FE estimator becomes inefficient and potentially biased. In contrast, the first difference estimator relaxes this assumption and remains consistent under weaker conditions, particularly when the error term follows a random walk or exhibits serial correlation. This probably make first difference a more robust choice for the data structure in this study.

Economically, it is also reasonable to assume that both train fares and flight demand follow a process resembling a random walk or at least exhibit serial correlation primarily through the random walk process. For example, train fares are typically adjusted gradually over time based on lagged cost drivers such as diesel prices or political decisions, rather than being completely reset each year. Similarly, flight demand is also shaped by long-run structural changes, including modal competition between transport modes or shifting consumer preferences. Forces like this tend to produce gradual trends rather than abrupt shifts, which means that the absolute levels of flight demand in a given year may reflect accumulated effects rather than immediate responses. Year-to-year changes in demand are therefore more likely to reflect short-term responses to pricing and service variations.

Moreover, with only 66 observations in the data set (6 routes over 12 years), there is a major risk that route fixed-effects would consume too much of the variation (Porath 2020). Including two-way fixed effects would be even more restrictive and further exacerbate this problem, as time fixed effects would absorb all variation in variables that only change over time (gasoline prices, GDP, OTP, and Covid-related shocks). This would leave too few degrees of freedom and eliminate the identifying variation needed to estimate the effects of interest.

4.2.3 Motivation and Intuition Behind the Covid Dummies and the Constraint

In the first difference estimator, the Covid dummy variable I from the levels model is split into two separate variables: $CovidB$, capturing the beginning of the pandemic in 2020, and $CovidE$, capturing its end in 2022. This adjustment is made to account for distinct time-specific demand shocks that affected travel behavior during the sample period. Including a single Covid dummy set to 1 throughout the pandemic would only allow variation to be identified when the dummy changes value (2020 and 2023). As a result, the model would not capture any effect during the intermediate years (2021 and 2022), despite ongoing disruptions related to the pandemic. By using two transition dummies, the model avoids this loss of identifying variation and is able to reflect both the initial drop in demand and the subsequent recovery. This structure provides a more accurate representation of the pandemic’s effect across its full duration.

The constraint $\beta_2 + \beta_3 = 0$ is imposed to align with the assumption of a symmetric Covid effect on air travel demand. In other words, the pandemic effect is assumed to be temporary and symmetric, remaining constant between the initial drop and the subsequent recovery. This restriction improves interpretability by allowing each coefficient to be seen as a deviation from the average pandemic effect, while also helping to reduce potential multicollinearity between time dummies. The constraint also enables a formal test of whether the entry and exit effects truly were symmetric, i.e., whether $\beta_2 = -\beta_3$.

4.3 Main Assumptions

As mentioned in Section 4.2.2, Motivation and Intuition Behind First Differences, the key assumption for the first difference estimator is that the change in error, Δu_{it}, between period t and $t-1$ follows a random walk. This means that the change in each regressor between the two time periods, Δx_{it}, is uncorrelated with the change in the error term (Wooldridge 2002). The identifying assumption is denoted:

$$E[(u_{it} - u_{i,t-1})(x_{it} - x_{i,t-1})] = 0$$

Additionally, the imposed constraint

$$\beta_2 + \beta_3 = 0$$

captures a symmetric Covid shock (i.e. an equal and opposite downturn and rebound). This restriction does not imply that the Covid-related effects are exactly assumed be identical in reality; rather, it provides a baseline for hypothesis testing.

5. Results

5.1 Main Estimation

This section presents the results from the first difference estimation of U.S. domestic flight demand, measured as $\Delta ln(Q^{\text{flight}}{}_{it})$. The model uses 60 first-differenced observations, derived from six routes over the years 2013-2023. The first year per route is lost due to differencing, resulting in ten observations per route. Standard errors are clustered at the route level.

Table 3 displays the results from three model specifications. The final (shaded) column represents the main result, based on the identification strategy outlined in Section 4.2.1, Econometric Specification. Column (1) presents the model without any Covid dummies. Column (2) adds the variables $CovidB$ (for 2020) and $CovidE$ (for 2022) and Column (3) imposes the linear restriction $\beta_2 + \beta_3 = 0$. This reflects the assumption that the pandemic had a constant effect on flight demand throughout the period it was active, with the full impact captured symmetrically at both the beginning and the end of the period.

Table 3. Results from the first difference estimation of domestic flight demand

	(1)	(2)	(3)
VARIABLES	$\Delta ln(Q^{\text{flight}}{}_{it})$	$\Delta ln(Q^{\text{flight}}{}_{it})$	$\Delta ln(Q^{\text{flight}}{}_{it})$
$\Delta ln(P^{\text{train}}{}_{it})$	1.399***	-0.368	0.203
	(0.217)	(0.291)	(0.277)
$\Delta ln(P^{\text{flight}}{}_{it})$	-1.024**	-0.828**	-0.943***
	(0.283)	(0.256)	(0.324)
$\Delta ln(P^{\text{gas}}{}_{t})$	0.315**	0.229	-0.116
	(0.107)	(0.145)	(0.0870)
$\Delta ln(GDP_t)$	43.15***	2.066	20.97***
	(2.371)	(4.569)	(2.975)
ΔOTP_t	0.0173***	-0.00234	0.00273
	(0.00397)	(0.00598)	(0.00504)
$CovidB$		-1.333***	-0.608***
		(0.177)	(0.0387)
$CovidE$		0.530***	0.608***
		(0.0262)	(0.0387)
$Constant$	-0.780***	0.0500	-0.374***
	(0.0426)	(0.105)	(0.0549)
Observations	60	60	60
R-squared	0.863	0.914	
Covid	NO	YES	YES
Constraint	NO	NO	YES

Standard errors, clustered at the route level, in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

5.2 Interpretation of Estimation Results

The cross-price elasticity of flight demand with respect to train fares is estimated at $e_{f,t} = 0.203$ in the main model (3), implying potential substitution between rail and air. However, the estimate is statistically insignificant and close to zero, providing no reliable evidence that higher train prices induce a meaningful shift to air travel. Moreover, the coefficient in column (1) is positive, high and statistically significant when Covid effects are omitted, but turns negative in column (2) when $CovidB$ and $CovidE$ are included without restriction. This instability in both sign and statistical significance suggests that the relationship between train fares and flight demand is not robust and likely sensitive to model specification. Therefore, no firm conclusion can be drawn regarding substitution patterns between flights and trains based on train pricing alone.

In contrast, the own-price elasticity of flight demand is estimated at $e_P = -0.943$, which is statistically significant at a 1% level. This suggests that a 1% increase in domestic air fares reduces passenger volumes by approximately 0.94%. Since $\mid e_P \mid = 0{,}94 < 1$, this indicates that flight demand is inelastic, meaning that quantity changes in demand are proportionally smaller than price changes (Nicholson & Snyder 2010:164). The estimated own-price elasticity remains negative, statistically significant, and consistent across all three specifications, which signal that the estimated price sensitivity of flight demand is robust to model variation.

The estimated effect of gasoline prices on domestic flight demand is inconsistent across specifications and does not yield a robust conclusion. In column (1), when Covid effects are not controlled for, the coefficient ($e_{f,g} = 0.315$) is positive and statistically significant at a 5% level. This would suggest that higher gasoline prices, making car travel more expensive, lead to increased demand for air travel. However, once Covid-related shocks are included in column (2) and (3), the estimates become smaller in magnitude and statistically insignificant. It even turns negative in the main model. This shift in both size and direction implies that the observed relationship between gasoline prices and the demand for domestic flights may be confounded by omitted variables and that the result is not reliable.

When it comes to the control variables, the estimated coefficient for GDP per capita shows a stable and significant positive association with flight demand in specification (1) and (3), indicating that higher income levels are linked to increased air travel. The estimate in specification (2) is also positive, but smaller and statistically insignificant, likely due to multicollinearity between GDP and Covid. The other control variable, Amtrak's on-time performance, is statistically insignificant and close to zero in the main estimation (3), indicating that train punctuality has negligible impact on flight volumes. The estimate in column (1), which excludes Covid controls, is on the other hand positive and statistically significant ($\beta = 0.0173$). This inconsistency may reflect either omitted variable bias in the simpler model or a structural break caused by the pandemic, which disrupted both travel behavior and service patterns.

Finally, the two COVID-19 dummies, $CovidB$ (2020) and $CovidE$ (2022), capture the collapse and rebound in flight demand associated with the pandemic. The coefficients in the restricted specification (3) are symmetric in magnitude but opposite in sign, reflecting the assumption that the pandemic's effect on demand was constant over the period it was active. The dummies are interpreted in a semilog framework, following Halvorsen & Palmquist (1980), which prescribes that a dummy coefficient c corresponds to a level change of $\exp(c) - 1$. At the onset of the pandemic in 2020 ($CovidB = 1$) the negative effect on flight demand is estimated to $\beta_2 = -0.608$. This implies a 0.608-point decrease in $\Delta ln(Q^{\text{flight}}{}_{it})$, which converts to $exp(-0{,}608) - 1 \approx -0.456$, i.e. a 45.6% decline in flight volume relative to the pre-Covid period. When the pandemic period is set to end in 2022 ($CovidE = 1$) the positive effect on flight demand is estimated to $\beta_3 = 0.608$. This yield $exp(0{,}608) - 1 \approx +0.837$, i.e. an 83.7% rebound from the depressed Covid level which restores flight volumes back to their pre-pandemic baseline. The coefficients on $CovidB$ and $CovidE$ are both highly statistically significant ($p < 0.01$), underscoring the pandemic's pronounced decline and subsequent rebound in flight demand. Although this symmetric specification fits the data well and aligns with the theoretical assumption of a constant pandemic effect, the robustness check in the next section will assess whether the effect truly remained stable throughout the pandemic or varied in magnitude between the collapse and the recovery.

5.3 Robustness and Alternative Specifications

In this section, the robustness of the Covid-related estimates is examined by relaxing the imposed symmetry constraint. The full regression results from this unrestricted specification are presented in Appendix C1. Additional robustness checks, including sequential adding of control variable to the first difference estimator and alternative model specifications are presented in Appendix B & C2-C5.

5.3.1 Robustness Test of the Symmetry Constraint

To check whether the imposed Covid constraint ($\beta_2 + \beta_3 = 0$) is supported by the data, I re-estimate the model replacing the $CovidB$ and $CovidE$ dummies with separate year dummies. Although the pandemic is defined as lasting from 2020 to 2022, I include dummies for 2020, 2021, 2022 and 2023. This is because the first difference estimator identifies effects through changes from year to year, meaning that recovery from the pandemic is captured in the transition between 2022 and 2023. Thus, a dummy for 2023 is required to capture this change.

Denote the dummies by C_0, C_1, C_2, C_3 respectively, then the dummy coefficients are defined by:

$\beta_{C_0} = C_0 - 0$, the level-shift in demand 2020 (onset of the pandemic).

$\beta_{C_1} = C_1 - C_0$, the change in the Covid-effect from 2020 to 2021.

$\beta_{C_2} = C_2 - C_1$, the change in the Covid-effect from 2021 to 2022.

$\beta_{C_3} = C_3 - C_2$, the level-shift from 2022 to 2023 (recovery from the pandemic).

Based on this, a series of Wald tests (Wooldridge 2013:818) is conducted to evaluate whether the assumption of a constant and symmetric Covid effect holds empirically.

Hypothesis HC1: Tests whether there is a change in the pandemic effect from 2020 to 2021. The null hypothesis is H_0: $\beta_{C_1} = 0$, which cannot be rejected ($p = 0.3682$), thus suggesting that the Covid effect remained stable between those years.

Hypothesis HC2: Tests whether the Covid effect was constant across the entire period from 2020 to 2022. Here, the null hypothesis H_0: $\beta_{C_2} + \beta_{C_1} = 0$ is rejected at a 5% level ($p = 0.0408$), indicating that the effect varied during the pandemic.

Hypothesis HC3: Tests whether the effect at the onset of the pandemic was symmetric to the recovery, i.e. the imposed constraint. The null hypothesis H_0: $\beta_{C_0} + \beta_{C_3} = 0$ is strongly rejected at a 1% level ($p = 0.0049$), providing evidence against the assumption of a symmetric Covid effect.

5.3.2 Justification for the Symmetry Constraint

The imposed constraint $\beta_2 + \beta_3 = 0$ is formally rejected in the Wald test corresponding to hypothesis HC3. This result suggests that the assumption of a perfectly symmetric Covid effect may be too restrictive, and that the imposed constraint does not fully align with the observed data. In contrast, the result from HC1 supports the idea of a stable Covid effect across the intermediate pandemic years providing some justification for treating the Covid effect as constant while the pandemic was ongoing. However, the fact that HC2 is rejected indicates that the cumulative impact of the pandemic evolved over time, implying that the assumption of a constant effect may be too restrictive beyond a single year. This likely reflects the influence of prolonged disruptions, changing policies and gradual adoptions in travel behavior.

Nonetheless, the imposed constraint still remains a theoretically motivated simplification that improves both precision and interpretability, by reflecting the idea of a temporary COVID-19 shock that was fully reversed at the end of the pandemic period.

6. Discussion

Although the climate benefits of shifting from air to rail are well documented, empirical research on the actual substitutability between these transport modes is still limited. This is concerning, since effective modal-shift policies require a clear understanding of the price relationships between transport modes (IPCC 2023). Previous studies, including Gama (2017), Escañuela Romana et al. (2023a), and Zeng et al. (2021), have found negligible or statistically insignificant cross-price elasticities between rail and air. This thesis contributes to the literature by presenting new estimates of cross-price elasticities based on annual U.S. route-level data from 2013 to 2023, capturing both long-run trends and the impact of the COVID-19 shock. Using a first difference estimator and relevant controls, the analysis isolates how year-to-year changes in train fares relate to changes in air travel demand. The following sections discuss the key findings, policy implications, study limitations, and directions for future research.

6.1 Analysis of Results and Implications

The main estimation reveals that the cross-price elasticity of flight demand with respect to train fares is small, positive, and statistically insignificant. This provides no strong evidence of substitution between long-distance rail and air travel in the U.S., at least not in response to price changes. The result is supported by the robustness checks, which show that both sign and statistical significance of the cross-price elasticity vary across model specifications (see Appendix B & C). Accordingly, Hypothesis **H1**: $e_{f,t} > 0$, as stated in Section 2.5, must be rejected as the result does not align with theoretical expectations regarding substitutable goods. However, the finding is consistent with earlier studies. For example, Gama (2017) reports an extremely low cross-price elasticity of 0.0008%, and Escañuela Romana et al. (2023a) similarly find weak and statistically insignificant cross-price elasticities between trains and flights in the U.S. These consistently small elasticities may reflect deeper structural barriers to modal substitution between rail and air, a point which is raised by Escañuela Romana et al. (2023a) and Wardman and Tyler (2000). The latter study emphasizes that fare levels, along with factors such as journey distance and prior service experience often play a decisive role in shaping travelers' mode choice. While this study includes rail on-time performance as a proxy for service quality, it does not account for journey distance, network accessibility, historical travel habits or other structural factors that may influence travel demand. This limitation is particularly important given that only around 1% of long-distance trips in the U.S. are made by train (Amtrak 2017). Rail is therefore clearly not the normative or expected mode of long-distance travel, which likely reduces the behavioral responsiveness to train fares. Many travelers may not perceive rail as a realistic substitute for either air or car travel, regardless of its price. The omission of these factors may partially account for the lack of statistical significance in the estimated cross-price elasticity, despite theoretical expectations of substitution between modes.

In contrast, the own-price elasticity of air travel demand is negative and statistically significant across all specifications, providing support for the Hypothesis **H2**: $e_P < 0$, stated in Section 2.5. The estimate from the main specification suggests that demand for domestic air travel price is slightly inelastic ($e_P = -0.94$), meaning that a 1% increase in air fares leads to a less proportional reduction in passenger volumes. This result is confirmed by robustness checks, which show that the own-price elasticity remains consistently negative and statistically significant across alternative specifications (see Appendix B & C). The result aligns with Escañuela Romana et al. (2023b), who estimate the own-price elasticity of demand for domestic flights on U.S routes to –0.70. Since this study is based on annual data, it is likely to reflect long-run behavioral adjustments. As Brons et al. (2002) emphasize, long-run elasticities are typically larger in magnitude than short-run elasticities, as consumers have more time to adapt to price changes and consider alternative travel options. Comparing the estimates and the statistical significance between the own- and cross-price elasticities suggests that direct pricing interventions (such as taxes or fees) targeted at air travel probably are more effective in reducing flight demand than indirect policies aimed at influencing train fares.

Gasoline prices were included to account for the cost of car travel. However, the magnitude of the effect varies notably across specifications and is not significant in the main model, leading to a rejection of Hypothesis **H3**: $e_{f,g} > 0$. In the absence of the constrained Covid dummies, the coefficient suggests a small positive relationship between gasoline prices and domestic flight demand. The lack of robustness indicates that the observed relationship may be affected by omitted variable bias and that no conclusions can be drawn regarding the role of gasoline prices in shaping air travel demand. This sensitivity is confirmed by the robustness checks, which show that both the sign and the statistical significance of the gasoline price coefficient vary across specifications (see Appendix B & C). Nonetheless, given that 90% of the long-distance trips in the U.S. are made by car, it remains relevant to control for car travel costs when analyzing long-distance travel in the country.

The finding that GDP per capita has a statistically significant and positive association with domestic air travel demand supports Hypothesis **H4**: $e_I > 0$, and is consistent with the interpretation of air travel as a normal or even luxury good

(Gallet & Doucouliagos, 2014). This suggests that rising income levels in the U.S. are likely to drive higher flight volumes and associated carbon emissions, unless any pricing or regulatory policies are imposed. When GDP is added to the first-difference model a positive relationship emerges directly (see Appendix B), but its effect appears less stable across alternative model specifications in the robustness checks (see Appendix C). This variation may result from distinctions in model structure, collinearity with Covid variables, or smaller sample sizes in some specifications. Nonetheless, it underlines the importance of considering income growth in climate policy design, as rising affluence may offset the impact of demand-reducing measures.

The lack of statistical significance for Amtrak's on-time performance (OTP) offers no support for Hypothesis **H5**: $\beta_1 < 0$. One possible explanation is that punctuality may not be a decisive factor in long-distance mode choice. It is possible that other unobserved service factors, such as travel time, comfort, or frequency (Wardman & Tyler 2000; Escañuela Romana et al. 2023a) play a more important role than punctuality alone.

Regarding the COVID-19 variables and the impact of the pandemic on travel demand, the results clearly demonstrate the importance of accounting for the pandemic when analyzing behavior during Covid-affected years. The statistically significant coefficients on the Covid dummies highlight the magnitude of the shock and its subsequent reversal in flight demand. Moreover, the robustness check (see Appendix C.1) reveals that while a simplified, symmetric specification offers a useful baseline, the actual effects of the pandemic were more complex and evolved over time. Such a pattern seems plausible, as people initially reacted strongly to restrictions at the onset of the pandemic, while health concerns and travel behavior gradually normalized as the situation progressed. This dynamic is also reflected in Column 2 of Table 3, where the estimated effects differ in magnitude between $CovidB$ and $CovidE$. These findings underscore the importance of properly modeling structural breaks, not only to improve the accuracy of parameter estimates, but also to avoid biased inference regarding key relationships in the model. Additional robustness tests (see Appendix C.3–C.5) further confirm the strong negative impact of the pandemic on flight demand across different model specifications.

6.2 Limitations

One major limitation of this study is the relatively small sample size. After excluding 2024 due to missing on-time performance data and accounting for the loss of the first year per route from differencing, only 60 observations remain.

This limits the statistical power of the analysis and makes the results more sensitive to model specification and the choice of control variables. Using annual data also prevents the identification of seasonal patterns and short-term fluctuations in demand and prices, which can be important for understanding travel behavior.

There are also several methodological limitations that need to be noted. Since the train fare proxy combines average passenger-related revenue per passenger mile with route-specific distance, it may not accurately reflect true price variations at the route level. This introduces potential measurement error (see Section 3., Data and Variable Construction for a detailed explanation of the proxy construction and its limitations). Another point is that while the first difference estimator helps address bias from unobserved time-invariant factors and mitigates endogeneity under the assumption of a random walk, endogeneity may still remain if the assumption does not hold. For example, if prices respond to short-term demand shocks not captured by the control variables, the differenced regressors may still correlate with the error term, leading to biased estimates. Furthermore, it's worth mentioning that although the random walk assumption is theoretically plausible, it is not empirically tested in this study.

There are also limitations related to the modeling of the COVID-19 pandemic. Robustness checks suggest that the pandemic's impact varied over time, which may bias the estimated effects. It also indicates the need for more flexible approaches to more accurately capture the dynamics of the pandemic period.

In addition, several potentially important explanatory variables are not included in the model due to data limitations and time constraints, despite their likely influence on travel mode choices. These include, among others, travel time, service frequency, accessibility of stations and airports, and comfort. Excluding such variables may lead to omitted variable bias.

Finally, even though this study focuses on a small number of domestic routes in the U.S., there are similarities to certain European routes in terms of travel distances and infrastructure. This means that the results could still be relevant for comparable contexts, particularly in Western Europe. That said, one should be cautious when applying the findings to places with very different transport systems, travel behavior, or policy environments.

6.3 Recommendations for future research

Several avenues for future research emerge from the limitations and findings of this study. To begin with, I recommend expanding the sample size (if the data availability permits). A larger dataset would improve the statistical power of the analysis and allow for identification of short-term fluctuations and seasonal patterns in travel behavior. Extending the geographical scope could also reveal important heterogeneities across regions and travel corridors.

This study focuses on the United States, where long-distance rail travel is limited. Future research could examine the cross-price elasticity between rail and air travel in countries with more established and widely used rail networks, such as those in Western Europe or East Asia. Comparative studies of this kind could improve the understanding how of substitution dynamics vary depending on levels in rail adoption and the overall development of transport system.

Given the profound and potentially lasting effects of the COVID-19 pandemic on travel behavior, further investigation into post-pandemic mobility patterns is essential. Understanding whether, and how, travel preferences have shifted permanently is crucial for designing effective transport policies in the years ahead.

Finally, since over 90% of long-distance trips in the U.S. are made by car, future research should explore the cross-price elasticity between car and rail travel in more depth. Such studies would offer a more comprehensive picture of intermodal competition and inform broader strategies for encouraging modal shift in the context of climate and congestion policy.

7. Conclusion

This thesis investigates the cross-price elasticity between rail and air travel in the United States, based on the research question: How do train ticket prices affect the demand for domestic flights in the U.S.? Using annual panel data from six routes over the period 2013–2023, the study applies a first difference estimator with a linear constraint to address endogeneity and time-invariant unobserved factors. The results indicate a small and positive, but statistically insignificant, cross-price elasticity between rail and air travel, providing no evidence of substitution between the two modes in the U.S. market. In contrast, the estimated own-price elasticity of domestic flight demand is statistically significant and inelastic (approximately -0.94), implying that air travel demand responds less than proportionally to fare increases. These findings suggest that policy efforts to encourage modal shift from air to rail will likely require more than adjustments in rail pricing alone. Direct pricing interventions on air travel, such as taxes or fees, may be more effective in reducing flight demand then policies aimed at lowering train fares. However, the analysis is limited by a small sample size, annual data, and a relatively narrow geographic focus, which reduces external validity. Additional limitations include simplified modeling of the COVID-19 period and the exclusion of potentially important explanatory variables like travel time and service frequency. Future research would benefit from richer datasets with higher temporal and spatial resolution, broader variable coverage, and a wider geographical scope. Comparative studies in countries with more established and widely used rail networks could offer further insights into how substitution patterns vary across transport systems and user contexts.

References

AAR (2025). *Freight Rail: Amtrak Support.* Association of American Railroads. https://www.aar.org/issue/freight-railroads-amtrak/ [2025-05-27]

Amtrak (2022). *CDP Climate Change 2022 Response*. Amtrak. https://www.amtrak.com/content/dam/projects/dotcom/english/public/documents/environmental1/CDP-2022.pdf [2025-05-21]

Amtrak (2024). *Fiscal Year 2023 Sustainability Report*. Amtrak. https://www.amtrak.com/content/dam/projects/dotcom/english/public/documents/environmental1/Amtrak-Sustainability-Report-FY23.pdf [2025-05-22]

Amtrak (2025a). *Reports & Documents*. https://www.amtrak.com/reports-documents [2025-04-29]

Amtrak (2025b). *Travel Guide to Train Fares - Easy Online Booking Options*. Amtrak. https://www.amtrak.com/guide-to-fares [2025-06-06]

Amtrak (n.d.). *CY 2023 Host Railroad Report Card & Route On-Time Performance*. https://www.amtrak.com/content/dam/projects/dotcom/english/public/documents/corporate/HostRailroadReports/Amtrak-2023-Host-Railroad-Report-Card.pdf. [2025-05-06]

Autor, D. (2016). *Lecture 4 - Utility Maximization*. https://ocw.mit.edu/courses/14-03-microeconomic-theory-and-public-policy-fall-2016/662896910b5530e160224afe6ac30752_MIT14_03F16_lec4.pdf [2025-05-26]

Berry, S., Levinsohn, J. & Pakes, A. (1995). Automobile Prices in Market Equilibrium. *Econometrica*, 63 (4), 841–890. https://doi.org/10.2307/2171802

Berry, S.T. (1994). Estimating Discrete-Choice Models of Product Differentiation. *The RAND Journal of Economics*, 25 (2), 242–262. https://doi.org/10.2307/2555829

Börjesson, M. (2014). Forecasting demand for high speed rail. *Transportation Research Part A: Policy and Practice*, 70, 81–92. https://doi.org/10.1016/j.tra.2014.10.010

Brons, M., Pels, E., Nijkamp, P. & Rietveld, P. (2002). Price elasticities of demand for passenger air travel: a meta-analysis. *Journal of Air Transport Management*, 8 (3), 165–175. https://doi.org/10.1016/S0969-6997(01)00050-3

Bureau of Transportation Statistics (2017). *National Household Travel Survey Long Distance Travel Quick Facts.* https://www.bts.gov/statistical-products/surveys/national-household-travel-survey-long-distance-travel-quick-facts [2025-05-20]

Bureau of Transportation Statistics (2024). *Hours of Delay and On-Time Performance of Amtrak: 2010-2023*. https://www.bts.gov/browse-statistical-products-and-data/info-gallery/hours-delay-and-time-performance-amtrak-2010-2023 [2025-05-06]

Drukker, D.M. (2003). Testing for Serial Correlation in Linear Panel-data Models. *The Stata Journal*, 3 (2), 168–177. https://doi.org/10.1177/1536867X0300300206

EIA (2025). *U.S. All Grades All Formulations Retail Gasoline Prices (Dollars per Gallon)*. EIA U.S Energy Information Administration. https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=EMM_EPM0_PTE_NUS_DPG&f=A [2025-04-29]

Escañuela Romana, I., Torres-Jiménez, M. & Carbonero-Ruz, M. (2023a). Elasticities of Passenger Transport Demand on US Intercity Routes: Impact on Public Policies for Sustainability. *Sustainability*, 15 (18), 14036. https://doi.org/10.3390/su151814036

Escañuela Romana, I., Torres-Jiménez, M. & Carbonero-Ruz, M. (2023b). Price Elasticity of Demand for Domestic Air Travel in the United States: A Robust Quasi-Experimental Estimation. *Atlantic Economic Journal*, 51 (2), 149–167. https://doi.org/10.1007/s11293-023-09779-4

Federal Railroad Administration (n.d.). *1.02 - Operational Data Tables*. https://safetydata.fra.dot.gov/OfficeofSafety/publicsite/Query/rrstab.aspx [2025-04-29]

FRED (2025). *Real gross domestic product per capita*. Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/A939RX0Q048SBEA [2025-04-29]

Gallet, C.A. & Doucouliagos, H. (2014). The income elasticity of air travel: A meta-analysis. *Annals of Tourism Research*, 49, 141–155. https://doi.org/10.1016/j.annals.2014.09.006

Gama, A. (2017). Own and cross-price elasticities of demand for domestic flights and intercity trains in the U.S. *Transportation Research Part D: Transport and Environment*, 54, 360–371. https://doi.org/10.1016/j.trd.2017.06.010

Halvorsen, R. & Palmquist, R. (1980). The Interpretation of Dummy Variables in Semilogarithmic Equations. *The American Economic Review*, 70 (3), 474–475

IER (2023). *New Registrations of Gasoline Vehicles Are Still Growing Despite the EV Push.* IER Institute for Energy Research. https://www.instituteforenergyresearch.org/fossil-fuels/gas-and-oil/new-registrations-of-gasoline-vehicles-are-still-growing-despite-the-ev-push/ [2025-04-29]

IPCC (ed.) (2023). Transport. In: *Climate Change 2022 - Mitigation of Climate Change.* 1. ed. Cambridge University Press. 1049–1160. https://doi.org/10.1017/9781009157926.012

Jevons, W.S. (1874). The Mathematical Theory of Political Economy. *Journal of the Statistical Society of London*, 37 (4), 478–488. https://doi.org/10.2307/2338697

Nicholson & Snyder (2010). *Microeconomic Theory: Basic Principles and Extensions*. 11th edition. Mason, OH: South-Western, Cengage Learning.

Porath, Y. (2020). *Sample size fixed effect regression?* ResearchGate. https://www.researchgate.net/post/Sample_size_fixed_effect_regression [2025-05-06]

Rail Passengers Association (n.d.). *Amtrak Ridership Statistics*. https://www.railpassengers.org/resources/ridership-statistics/ [2025-04-29]

Ritchie, H. (2023). *Which form of transport has the smallest carbon footprint?* Our World in Data. https://ourworldindata.org/travel-carbon-footprint [2025-05-21]

Ritchie, H., Rosado, P. & Roser, M. (2024). *Greenhouse gas emissions.* Our World in Data. https://ourworldindata.org/greenhouse-gas-emissions [2025-05-04]

Rohr, C., Fox, J., Daly, A., Patruni, B., Patil, S. & Tsang, F. (2013). Modeling Long-Distance Travel in Great Britain. *Transportation Research Record*, 2344 (1), 144–151. https://doi.org/10.3141/2344-16

Slutsky, E. (1915). Sulla teoria del bilancio del consumatore. *Giornale degli Economisti e Rivista di Statistica*, 51 (Anno 26) (1), 1–26

U.S. Department of Transportation (n.d.). *Travel Time Reliability*. https://ops.fhwa.dot.gov/publications/tt_reliability/brochure/ttr_brochure.pdf [2025-05-06]

U.S. Department of Transportation (2025). *Consumer Airfare Report: Table 6 - Contiguous State City-Pair Markets That Average At Least 10 Passengers Per Day*. Department of Transportation - Data Portal. https://data.transportation.gov/Aviation/Consumer-Airfare-Report-Table-6-Contiguous-State-C/yj5y-b2ir/about_data [2025-04-29]

US EPA, O. (2025). *Carbon Pollution from Transportation*. [Overviews and Factsheets]. https://www.epa.gov/transportation-air-pollution-and-climate-change/carbon-pollution-transportation [2025-05-20]

Varian, H.R. (2010). *Intermediate microeconomics: a modern approach*. 8. ed. W.W. Norton & Company.

Wardman, M. & Tyler, J. (2000). Rail network accessibility and the demand for inter-urban rail travel. *Transport Reviews*, 20 (1), 3–24. https://doi.org/10.1080/014416400295310

Wooldridge, J.M. (2002). *Econometric Analysis of Cross Section and Panel Data*. 2nd edition. The MIT Press Cambridge.

Wooldridge, J.M. (2013). *Introductory Econometrics: A Modern Approach*. 5th Edition. Mason, OH: South-Western, Cengage Learning.

World Health Organization (2023). *From emergency response to long-term COVID-19 disease management: sustaining gains made during the COVID-19 pandemic*. https://www.who.int/publications/i/item/WHO-WHE-SPP-2023.1 [2025-05-12]

Zeng, Y., Ran, B., Zhang, N. & Yang, X. (2021). Estimating the Price Elasticity of Train Travel Demand and Its Variation Rules and Application in Energy Used and CO2 Emissions. *Sustainability*, 13 (2), 475. https://doi.org/10.3390/su13020475

Appendix A – Correlation Matrix

Based on the correlation matrix in Table 4, the strongest relationship is observed between flight demand and flight prices, with a negative correlation of -0.582 ($p < 0.01$). This result is expected, as it aligns the economic theory and supports the Hypothesis **H2**: $e_P < 0$, outlined in Section 2.5, Hypotheses.

The correlation between train prices and flight demand is positive and statistically significant at 0.429. This is consistent with the theoretical expectation of substitutable transport modes and supports Hypothesis **H1**: $e_{f,t} > 0$. However, it is important to note that correlation does not imply anything about causality, meaning that the relationship should be interpreted with caution.

In contrast, the correlation between flight demand and gasoline prices is weak and statistically insignificant, representing a small or no direct association. Surprisingly the correlation between flight demand and GDP per capita also is insignificant, indicating that other factors than income level in the U.S. probably have a more prominent role in explaining variation in air travel demand. The impact from on-time performance appears to be negligible.

Table 4. Pairwise correlation matrix

Variables	(1)	(2)	(3)	(4)	(5)	(6)
(1) Q_flight	1.000					
(2) P_train	0.429	1.000				
	(0.000)					
(3) P_flight	-0.582	-0.451	1.000			
	(0.000)	(0.000)				
(4) P_gas	0.008	-0.164	0.097	1.000		
	(0.946)	(0.169)	(0.417)			
(5) GDP	-0.044	-0.031	0.060	0.402	1.000	
	(0.713)	(0.794)	(0.619)	(0.000)		
(6) OTP	-0.136	0.014	-0.116	-0.121	-0.204	1.000
	(0.275)	(0.914)	(0.355)	(0.332)	(0.101)	

Appendix B - Sequential Adding of Control Variables

In this section I examine how the coefficients of interests evolve as additional covariates sequentially are introduced one by one. Table 5 reports seven separate first difference regressions of $\Delta ln(Q^{\text{flight}}{}_{it})$, each column adding the next variable in the sequence.

Table 5. Sequential adding of control variables in the first difference estimator

VARIABLES	(1) $\Delta ln(Q^{\text{flight}}{}_{it})$	(2) $\Delta ln(Q^{\text{flight}}{}_{it})$	(3) $\Delta ln(Q^{\text{flight}}{}_{it})$	(4) $\Delta ln(Q^{\text{flight}}{}_{it})$	(5) $\Delta ln(Q^{\text{flight}}{}_{it})$	(6) $\Delta ln(Q^{\text{flight}}{}_{it})$	(7) $\Delta ln(Q^{\text{flight}}{}_{it})$
$\Delta ln(P^{\text{train}}{}_{it})$	-0.833***	-0.897***	-0.587**	1.180***	1.399***	-0.368	0.203
	(0.172)	(0.211)	(0.176)	(0.221)	(0.217)	(0.291)	(0.277)
$\Delta ln(P^{\text{flight}}{}_{it})$		0.669	0.341	-0.952**	-1.024**	-0.828**	-0.943***
		(0.422)	(0.432)	(0.342)	(0.283)	(0.256)	(0.324)
$\Delta ln(P^{\text{gas}}{}_{t})$			1.141***	0.188	0.315**	0.229	-0.116
			(0.105)	(0.109)	(0.107)	(0.145)	(0.0870)
$\Delta ln(GDP_t)$				39.26***	43.15***	2.066	20.97***
				(3.140)	(2.371)	(4.569)	(2.975)
ΔOTP_t					0.0173***	-0.00234	0.00273
					(0.00397)	(0.00598)	(0.00504)
$CovidB$						-1.333***	-0.608***
						(0.177)	(0.0387)
$CovidE$						0.530***	0.608***
						(0.0262)	(0.0387)
$Constant$	0.0210	0.0159	0.0191	-0.705***	-0.780***	0.0500	-0.374***
	(0.0154)	(0.0181)	(0.0177)	(0.0507)	(0.0426)	(0.105)	(0.0549)
Observations	66	66	66	66	60	60	60
R-squared	0.224	0.255	0.380	0.838	0.863	0.914	
Covid	NO	NO	NO	NO	NO	YES	YES
Constraint	NO	NO	NO	NO	NO	NO	YES

Standard errors, clustered at the route level, in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

In the first three specifications, the estimated cross-price elasticity of air travel demand with respect to train fares is significantly negative, surprisingly implying that higher train prices reduce flight demand. Also, the own-price elasticity of air fares is counterintuitive since it appears positive. Once the control $\Delta ln(GDP_t)$ is introduced in (4), both elasticities flip to their theoretically expected signs. Adding ΔOTP_t in specification (5) leaves the other coefficients stable, suggestion that fluctuations in train punctuality are uncorrelated with the elasticities. In (6), the inclusion of $CovidB$ and $CovidE$ dummies make the cross-price elasticity insignificant, while the own-price elasticity of flight demand remains negative and significant. This suggests that the Covid dummies capture much of the variation that was previously attributed to train fares. Specification (6) and (7) estimate the same underlying first difference model of $\Delta ln(Q^{\text{flight}}{}_{\text{it}})$, but differ in how the Covid shocks are handled. In (6) the Covid-coefficients β_2 and β_3 are left unrestricted, allowing for an asymmetric the collapse ($\beta_2 = -1.333$) and rebound ($\beta_3 = 0.530$). The last specification (7) presents the results from the main estimation, incorporating all controls and the impose symmetry constraint $\beta_2 + \beta_3 = 0$.

Appendix C - Robustness and Alternative Specifications

C.1 Results from the Unrestricted Specification with Separate Year Dummies

In Table 6 are the results from the unrestricted specification, which includes four year dummies for 2020, 2021, 2022, and 2023. These estimates are used for the Wald tests in Section 5.3.1, Robustness Test of the Symmetry Constraint.

Table 6. Results from the unrestricted first differences estimator

VARIABLES	$\Delta ln(Q^{\text{flight}}{}_{it})$
$\Delta ln(P^{\text{train}}{}_{it})$	-0.0801
	(0.410)
$\Delta ln(P^{\text{flight}}{}_{it})$	-0.848**
	(0.220)
$\Delta ln(P^{\text{gas}}{}_{t})$	0.260
	(0.148)
$\Delta ln(GDP_t)$	-3.594
	(4.649)
ΔOTP_t	-0.00277
	(0.00535)
C_0	-1.636***
	(0.311)
C_1	0.486
	(0.492)
C_2	0.456**
	(0.154)
C_3	0.145***
	(0.0285)
$Constant$	0.122
	(0.0963)
Observations	60
R-squared	0.921

Standard errors, clustered at the route level, in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The separate year dummies capture how domestic flight demand changed from one year to the next during the COVID-19 period. The dummy for 2020 (C_0) represents the initial collapse in demand between 2019 and 2020 and is large and statistically significant at a 1% level, indicating an approximately 80.5% decline in flight passengers. The dummy for 2021 (C_1) is not statistically significant, suggesting that demand remained relatively stable between 2020 and 2021. On the other hand, the coefficient for 2022 (C_2) is positive and statistically significant at a 5% level, corresponding to a 57.8% increase in domestic flight demand between 2021 to 2022. Finally C_3 captures the continued rebound from 2022 to 2023 and shows a statistically significant 15.6% increase. This result suggest that the pandemic's impact evolved over time, with a sharp decline in 2020 followed by a gradual recovery.

C.2 Estimation on Pre-Pandemic Data

As another robustness check, the first-difference estimator is restricted to the pre-pandemic period (2013–2019). In practice, this means that the regressions only use observations with Year ≥ 2013 and ≤ 2019, ensuring that the estimate of the cross-price elasticity is not driven by the COVID-19 years. This reduces the dataset to only 36 observations. The results from the estimation are presented in Table 7.

Table 7. Results from the first difference estimator on pre-pandemic data

VARIABLES	$\Delta ln(Q^{\text{flight}}{}_{it})$
$\Delta ln(P^{\text{train}}{}_{it})$	-0.229
	(0.401)
$\Delta ln(P^{\text{flight}}{}_{it})$	-0.727***
	(0.0957)
$\Delta ln(P^{\text{gas}}{}_{t})$	0.256
	(0.132)
$\Delta ln(GDP_t)$	-5.231
	(7.273)
ΔOTP_t	-0.00398
	(0.00688)
$Constant$	0.156
	(0.148)
Observations	36
R-squared	0.401

Standard errors, clustered at the route level, in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The estimated cross-price elasticity between rail and air turns negative, but remains statistically insignificant. This indicates no clear substitution effect even before the COVID-19 shock based on the limited data set. The own-price elasticity of flight demand remains negative, statistically significant and relatively similar in magnitude (-0.73), supporting the robustness of the finding that air travel demand is inelastic. All other coefficients lose significance, likely due to the reduced number of observations, which increases standard errors and lowers the statistical power of the test.

C.3 Single Fixed Effects

An alternative to a first difference estimator is a single fixed effects (FE) estimator, which controls for unobserved route specific time-invariant heterogeneity (α_i). The fixed effects model retains the same functional form as the levels specification introduced in Section 4.1.1:

$$ln(Q^{\text{flight}}{}_{it}) = \beta_0 + e_{f,t}\, ln(P^{\text{train}}{}_{it}\,) + e_P\, ln(P^{\text{flight}}{}_{it}) + e_{f,g}\, ln(P^{\text{gas}}{}_{t}) \\ + e_I\, ln(GDP_t) + \beta_1 OTP_t + C \cdot I(Covid\ =\ 1) + \alpha_i + u_{it}$$

The fixed effects estimator relies on the assumption of strict exogeneity, meaning that the explanatory variables are uncorrelated with the idiosyncratic error term in all time periods, conditional on the unobserved effect:

$$E(u_{it}|x_i, c_i) = 0 \qquad \text{(Wooldridge 2002:266).}$$

The fixed effects approach (unlike random effects) allows the unobserved, time-invariant characteristics to be arbitrarily correlated with the explanatory variables. This flexibility makes the estimator robust to omitted variable bias arising from unobserved, constant factors that may influence both prices and demand.

Table 8 on the next page reports the results from the single fixed effects estimator. The coefficient on train fares is negative and marginally significant, which contrasts with stated hypothesis **H1**: $e_{f,t} > 0$ from Section 2.5 and the positive, though insignificant, estimate in the first difference model. This provides support for that the relationship between train fares and flight demand is not robust and probably sensitive to both model and underlying assumptions. The own-price elasticity of flight demand ($e_P \approx -0.75$) confirms earlier findings that air travel is price inelastic.

Table 8. Results from the single fixed effects estimator

VARIABLES	$ln(Q^{\text{flight}}{}_{it})$
$ln(P^{\text{train}}{}_{it})$	-0.423*
	(0.177)
$\ln(P^{\text{flight}}{}_{it})$	-0.752*
	(0.296)
$ln(P^{\text{gas}}{}_{t})$	0.271**
	(0.0990)
$ln(GDP_t)$	0.911
	(0.665)
OTP_t	-0.0285***
	(0.00573)
$Covid$	-0.825***
	(0.0323)
$Constant$	9.425
	(6.618)
Observations	66
Number of route_id	6
R-squared	0.723

Standard errors, clustered at the route level, in parentheses
*** p<0.01, ** p<0.05, * p<0.1

C.4 Two-Way Fixed Effects

As an another robustness test the demand specification in levels is estimated including two-way fixed effects. Concretely this model is estimated:

$$
\begin{aligned}
ln(Q^{\text{flight}}{}_{it}) = \beta_0 + e_{f,t}\, ln(P^{\text{train}}{}_{it}) + e_P\, ln(P^{\text{flight}}{}_{it}) + e_{f,g}\, ln(P^{\text{gas}}{}_{t}) \\
+\ e_I\, ln(GDP_t) + \beta_1 OTP_t\ +\ C \cdot I(Covid = 1) + \alpha_i + \gamma_t + u_{it}
\end{aligned}
$$

where α_{i} represents all time-invariant characteristics of each route, and γ_{t} captures all year-specific, time-varying shocks that affect all routes equally, e.g. macroeconomic changes and the COVID-19 pandemic.

When estimating the levels model with both route- and time fixed effects, five years (2019-2023) are omitted due to collinearity problems, see Table 9. The high significance of the estimates in the levels model probably reflects that the two-way fixed effects absorb a large share of the variation, leaving less unexplained variance and thereby producing smaller standard errors. This can result in statistically significant estimates even in relatively small samples, although it also increases the risk of overfitting. This contrasts with the first-difference specification using only pre-pandemic data, where the reduced sample size and limited variation lead to larger standard errors and less statistical power.

Table 9. Results from the two-way fixed effects estimator

VARIABLES	$ln(Q^{\text{flight}}{}_{it})$
$ln(P^{\text{train}}{}_{it})$	-0.620**
	(0.181)
$\ln(P^{\text{flight}}{}_{it})$	-0.758*
	(0.304)
$ln(P^{\text{gas}}{}_{t})$	0.432***
	(0.0834)
$ln(GDP_t)$	-8.830***
	(0.471)
OTP_t	-0.206***
	(0.0134)
$Covid$	-0.423***
	(0.0280)
2014.Year	-1.724***
	(0.0884)
2015.Year	-1.657***
	(0.122)
2016.Year	0.176**
	(0.0458)
2017.Year	-0.541***
	(0.0706)
2018.Year	-0.570***
	(0.0890)
2019o.Year	-
2020o.Year	-
2021o.Year	-
2022o.Year	-
2023o.Year	-
Constant	131.3***
	(5.425)
Observations	66
Number of route_id	6
R-squared	0.900

Standard errors, clustered at the route level, in parentheses
*** p<0.01, ** p<0.05, * p<0.1

C.5 Random Effects

As a complementary robustness test, the same demand specification is also estimated using a random effects estimator:

$$ln(Q^{\text{flight}}{}_{it}) = \beta_0 + e_{f,t}\, ln(P^{\text{train}}{}_{it}) + e_P\, ln(P^{\text{flight}}{}_{it}) + e_{f,g}\, ln(P^{\text{gas}}{}_{t}) + e_I\, ln(GDP_t) + \beta_1 OTP_t + C \cdot I(Covid = 1) + u_{it}$$

Unlike fixed effects, the random effects approach assumes that unobserved, time-invariant route characteristics are uncorrelated with the explanatory variables. Formally, this requires the orthogonality condition

$$E(\alpha_i|x_i) = E(\alpha_i) = 0$$

where α_i denotes the unobserved route-specific effect and $x_i = (x_{i1}, x_{i2}, \dots, x_{iT})$ represents the explanatory variables for route i across time (Wooldridge 2002:257). This assumption allows the model to exploit both within- and between-route variation, potentially improving efficiency compared to fixed effects.

However, this assumption is likely violated in the present context, as unobserved factors such as underlying service quality or persistent patterns in traveler demand, like a high share of business travelers, may influence both prices and demand. Therefore, while random effects is not suitable as the main estimation method, it serves as a useful robustness check. The results remain consistent in terms of significance for the train and flight fare, supporting the stability of the main findings.

Table 10. Results from the random effects estimator

VARIABLES	$ln(Q^{\text{flight}}{}_{it})$
$ln(P^{\text{train}}{}_{it})$	-0.0333
	(0.146)
$ln(P^{\text{flight}}{}_{it})$	-1.048***
	(0.273)
$ln(P^{\text{gas}}{}_{t})$	0.511***
	(0.137)
$ln(GDP_t)$	0.866
	(0.708)
OTP_t	-0.0268***
	(0.00457)
$Covid$	-0.854***
	(0.0449)
$Constant$	9.242
	(7.868)
Observations	66
Number of route_id	6

Standard errors, clustered at the route level, in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Publishing and archiving

Approved students' theses at SLU can be published online. As a student you own the copyright to your work and in such cases, you need to approve the publication. In connection with your approval of publication, SLU will process your personal data (name) to make the work searchable on the internet. You can revoke your consent at any time by contacting the library.

Even if you choose not to publish the work or if you revoke your approval, the thesis will be archived digitally according to archive legislation.

You will find links to SLU's publication agreement and SLU's processing of personal data and your rights on this page:

- https://libanswers.slu.se/en/faq/228318

☒ YES, I, Stina Andersson, have read and agree to the agreement for publication and the personal data processing that takes place in connection with this

☐ NO, I/we do not give my/our permission to publish the full text of this work. However, the work will be uploaded for archiving and the metadata and summary will be visible and searchable.