DISSERTATION

ESSAYS ON AVIATION, REGIONAL ECONOMIC DEVELOPMENT, AND MIGRATION

Submitted by

Ulmaskhon Kalandarova

Department of Economics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2025

Doctoral Committee:

Advisor: Anita Alves Pena
Co-Advisor: Stephan Weiler

Daniele Tavani
Gregory Graff

ABSTRACT

ESSAYS ON AVIATION, REGIONAL ECONOMIC DEVELOPMENT, AND MIGRATION

Aviation is an important factor of promoting regional development. Its impact to the economy could be analyzed from various angles, including through the effect of airports to local and regional economies, and through the effects of novel aviation fuels to the local economies and sustainability, in general. The first two essays in this dissertation examine the effect of aviation from those above-mentioned angles to local economies in Colorado and California. The third essay focuses on the phenomenon of migration and remittances during the post-Soviet era in Tajikistan, a country in Central Asia.

Motivated by reverse causality issue between airport development and regional economic growth, in Chapter 1, titled as *"Air Transportation and Regional Economic Development in Colorado"*, we examine the dynamic causal relationship between air transportation at Denver International Airport (DIA) and regional economic development for the Denver MSA and the overall state of Colorado using time series methods for the period from 2000 to 2019. The results reveal overall positive effects of air transportation on local regional development in the Denver MSA with muted effects for the state of Colorado as a whole, consistent with impacts fading for more remote areas. Estimation of VECM models and Impulse Response Function reveals bi-directional, positive, significant long-run causality between Total Domestic Flight Passengers and Total Business Entities in Denver: a 1% change in Total Domestic Flights Passengers leads to a 3% increase in Total Registered Businesses and this magnitude is higher than in the vice-versa case. Total Domestic Flight Passengers are also positively associated with local employment in

Denver, in contrast to a negative effect for the larger Colorado region in the short-run. A 1% change in Total Domestic Flights Passengers leads to a 0.006% increase in Total Non-Farm Employment in Denver and about a 0.004% decrease in Colorado broadly. Total Cargo affects employment positively both in Denver and beyond, estimated by both VAR model and Impulse Response Functions. The same 1% change in Total Cargo is associated with about a 0.007% increase in Total Non-Farm Employment in Denver and about a 0.005% increase in Colorado. The paper confirms traditional claims that airport services play an important role in local and regional development measured via both traditional employment and non-traditional business variables. Furthermore, comparison of the Denver MSA to the wider state illustrates differences associated to the distance to the hub city.

Chapter 2, titled as “Inter-fuel Substitution in the Aviation Sector and Sustainable Aviation Fuel: Case study of California”, explores the dynamics of inter-fuel substitution, focusing on the impact of biofuels in the aviation sector and their potential in decarbonizing aviation in California. Specifically, the research investigates two key objectives: (1) understanding the historical interplay of substitution or complementary relationships among general transportation fuels after the mid-2010’s introduction of biofuels, and (2) evaluating the extent to which biofuel adoption drives inter-fuel substitution in particularly aviation sector. Ordinary Least Squares (OLS) and Seemingly Unrelated Regression (SUR) methods are used to analyze the data from U.S. Energy Information Administration (EIA) and U.S. Environmental Protection Agency (EPA) for 1980-2022 to provide a comprehensive view of biofuel integration. Results reveal that while biofuels have not significantly changed traditional substitution and complementarity relationships in the general transportation sector, they serve as a viable alternative to kerosene-based jet fuel in aviation. A 1% increase in jet fuel price leads to 1.043% increase of SAF volumes consumed, demonstrating that

SAF has become a valid substitute for a traditional kerosine-based jet fuel. The effect of price changes of SAF on jet fuel volumes still remains limited because of small shares of biofuels in aviation. The paper demonstrates biofuels' growing potential to reduce dependency on fossil fuels and their important contribution to achieving long-term sustainability in aviation.

Finally, in Chapter 3, titled as “Motives of Migration and Remittances in Tajikistan”, we focus on migration phenomena in Tajikistan, analyzing what factors drive people to migrate from Tajikistan and to remit back to Tajikistan. Probit and Heckman sample selection correction methods are used to analyze migrant and household characteristics data from the World Bank’s “Jobs, Skills, and Migration Survey in Tajikistan, 2013.” Results reveal that migration cost variables like having networks and family arrangement of costs related to migration have a highest strong positive association with the probability of having migrants in the households, empirically reinforcing the evidence of how important migration networks and family arranged costs of migration are in shaping the migrations flows from Tajikistan to Russia, making diasporas important “non-formal” institutions between two countries. The college education and language skills variables reveal a negative association with the probability of having migrants in the households, complying with a Negative Selection migration process predicted by Roy’s model (1951). Language skills demonstrate a positive association with the amount of remittances, since it makes easier to find a higher-paid job knowing the destination language, and so more remittances could be sent to Tajikistan. Also, family size and family income show a positive association with remittances transferred to households. Additionally, the results reveal that there is a significant selection bias (errors from two equations are correlated, lambda is negative and significant) and therefore Heckman estimation method usage is justified to correct for the self-selection bias.

ACKNOWLEDGEMENTS

I would like to express my deeply, sincere and heartful gratitude to my advisors Dr. Anita Pena and Dr. Stephan Weiler – I cannot be enough thankful to you for all the support, guidance, and for always helping and caring for me in each and every situation. Without you I wouldn't be able to made it! Also, I am very grateful to my Dissertation Committee, Dr. Daniele Tavani and Dr. Gregory Graff, who also supported me greatly in this journey! Thank you All so much! I am honored to be your student and to have such amazing Advisors as You All! – I wouldn't have made it without All of you!

I am also very thankful to my amazing cohort and friends: Manuel Cruz, Wisnu Nugroho, Ashish Sedai, Michelan Willson, Sayorn Chin, Fatih Kirsanli, Adam Walke, Sanchari Choudhury – who inspired me and shared all life in graduate school together.

Lastly, my Family – my Parents and Sobir aka – all your love, patience, sacrifice could not be compensated enough by me. Thank God and to You All!

DEDICATION

For my Beloved Family: my Dad, my Mom and Sobir aka.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter-1:
"Introduction: Aviation, Migration and Regional Economic Development"

My research interest lies in the fields of aviation and regional economic development. So, in the first part of this dissertation we tried to explore particularly the economic impact of airports on local communities. While there is a substantial body of literature exploring how airports contribute to local economic development, much of it focuses on the one-way relationship—how airport activity stimulates growth in surrounding areas. However, less attention has been given to the reverse causality, where local economic conditions may influence airport activity. This reverse dynamic is of particular interest in the case of Denver International Airport (DIA), where we aim to explore whether it is the airport driving economic growth in Denver or whether the city's expanding economy is fueling greater airport activity. Moreover, while DIA serves Denver, it is also a crucial hub for the entire state of Colorado, meaning that its economic impact stretches beyond the immediate vicinity. This broader regional perspective motivates us to investigate how airport activity affects not just Denver but also more remote areas of Colorado. To explore these dynamics, we use traditional variables like air traffic and employment figures, but also incorporate less conventional metrics such as the number of business entities and net cargo volume, to capture a more comprehensive view of how airport activity influences economic development across the region. By addressing both the direct and indirect effects of airport activity, we aim to provide a clearer understanding of the complex relationship between aviation and local/regional economies.

In the second part of this dissertation, we shift our focus to the role of novel biofuels, specifically sustainable aviation fuels, in reducing CO_2 emissions from aviation. They are a new type of biofuel that has recently been introduced to the aviation fuel market. Our goal is to examine the extent to which biofuels have become a competitive substitute for traditional jet fuels. We concentrate our analysis on California, as it is the primary state where most SAF production and imports are consumed, particularly at its airports. California also stands out as a leader in the U.S. when it comes to implementing advanced programs and strategies aimed at transitioning to a more sustainable, green economy, including in the aviation sector. Given California's prominent role in promoting environmentally friendly initiatives, investigating the potential inter-substitutability between biofuel-based SAFs and conventional jet fuels is crucial for evaluating the effectiveness of the state's efforts to support the aviation industry's shift toward more greener practices. This analysis provides valuable insights into how well SAFs are positioned to contribute to California's broader environmental goals and existing aviation fuels market in particular.

In the third part of this dissertation, we explore a topic distinct from aviation: the phenomenon of migration from Tajikistan, a post-Soviet country in Central Asia. Migration has become an increasingly significant issue for Tajikistan over the last 25-30 years, largely due to the economic challenges the country has faced since gaining independence in 1991. Tajikistan, one of the poorest countries in the region, has experienced high levels of emigration as a result of limited employment opportunities, low wages, and economic instability. A large proportion of the population has sought work abroad, particularly in Russia, to send remittances back home. These remittances have become a critical source of income for many households and are an essential part of the Tajik economy. And there is still a lack of empirical literature in the field of motives of migration in Tajikistan.

Chapter-2:
"Air Transportation and Regional Economic Development in Colorado"

2.1. Introduction

Airports play an important role in stimulating local economies, acting as central hubs for trade, tourism, and job creation. They provide direct employment opportunities through roles in airport operations, airlines, and security. Airports also generate indirect jobs in sectors such as tourism, hospitality, and retail. Airports have often been seen as potential main facilitators for economic development, and several studies revealed evidence of their importance in the development of some high-technology centers, such as Silicon Valley in California, Route 128 in Boston, and the M4 Corridor outside of London (Button et al., 2009).

Studies focusing on international airports, particularly in the U.S., indicate that their presence fosters substantial economic benefits in nearby areas. Service-oriented industries and high-tech sectors, in particular, experience considerable gains from proximity to airports, which improve connectivity and access to global markets (Appold, 2015; Appold & Kasarda, 2013; Brueckner, 2003; Button & Taylor, 2000; Cidell, 2014; Sheard, 2014). Lenaerts et al. (2021) proposes several economic mechanisms, through which air transportation effects the regions: businesses will prefer to settle in the areas with better access to air transportation, because it will decrease their transportation costs. These population concentrations are then related to the development of increased amenities that attract more people and businesses to the regions.

The majority of papers in the literature that examine the effect of air transportation on regional development focus on regional employment and regional income variables. Fewer studies examine other regional development indicators like emerging businesses in response to air transportation development (The Economic Impact of Civil Aviation on the U.S. Economy, Federal Aviation Administration, 2016).

Furthermore, although numerous studies suggest positive spillovers of airports on local economies, the extent of reverse causality remains elusive. Reverse causality between airport development and regional economic growth refers to the challenge of determining whether the construction and expansion of airports lead to regional development or if existing economic growth in the region spurs the development of airports. In this relationship, it is unclear whether improved airport infrastructure acts as a catalyst for local growth or if regions that are already economically thriving invest in airports to support their expansion. Numerous studies have explored this dynamic, highlighting the difficulty in establishing a clear direction of causality due to pre-existing economic strength in the regions (Sheard, 2019; Button & Lall, 1999; Bel & Fageda, 2008; Blonigen & Cristea, 2015). Since airports and economic growth may play mutually reinforcing roles, it is difficult to pinpoint a singular direction of causality.

Another complexity surrounds endogeneity between the experiences with and across regions. Literature that focuses on a particular single region and its causal relationship with existing airport close to that region (e.g., Button et al., 2009) is an accurate approach in places where airports are located within geographically smaller service areas. The case study presented in this paper, however, illustrates a possible counter-example. Denver International Airport (DIA) has a vast air service network and is ranked as the 3rd busiest airport in the U.S. (Metro Denver EDC, 2023). It is unlikely that DIA provides services just to the nearby Denver Metropolitan Statistical Area (MSA). Instead, DIA is the only major hub airport in the State of Colorado, so people who want to travel from other places to remote areas inside Colorado still use DIA as their main flight hub point (Colorado Department of Transportation, 2018).

The purpose of this paper is to analyze the effect of air traffic at Denver International Airport on the emergence and growth of businesses in the Denver MSA area and change of employment in both the Denver MSA and Colorado more broadly. This case study provides a setting to more precisely pinpoint any reverse causality between airports and local economies using both traditional (employment and/or income) and non-traditional (emerging businesses) variables and to compare city-hub versus periphery settings.

Based on time series methods (ARDL, VAR, VECM, Granger Causality, Impulse Response Functions), we find an overall positive effect of air transportation on local regional development in Denver metro area that fades out when more remote areas in Colorado are included. This is intuitive if business concentration is mostly located inside or close to the Denver metro area. The direction of reverse causality is stronger from airport to regional development rather than vice-versa, which reinforces the idea of higher importance of airports to local economies than local economies on airport development.

The paper is organized as follows. The next section summarizes the existing literature about effects between air transportation and local economic development, emphasizing endogeneity issues between local and neighboring regions. Subsequently, the data set and estimation strategy are outlined. The results section presents the main results and discussion. The last section concludes.

2.2. Literature Review and Contextual Background

Airports are widely recognized as essential infrastructure that can significantly contribute to regional development by facilitating economic growth, enhancing accessibility, and promoting social integration. The impact of airports on regional development can be examined through various aspects, such as economic activity, employment generation, business development, tourism, and investment attraction. A broad strand of the literature attributes positive regional spillovers to airports.

2.2.1. Economic Activity and Growth, Employment Creation

Airports serve as catalysts for economic growth by enhancing connectivity, which facilitates trade and the movement of goods and services. Research indicates that regions with developed airport infrastructure experience higher economic activity and growth compared to those without (Green, 2007). Airports support air transport-dependent industries, such as manufacturing and technology, which require quick supply chains and global market access. For instance, Button et al. (2009) highlight the role of airports in driving growth in high-tech areas like Silicon Valley, Boston's Route 128, and London's M4 Corridor by attracting businesses reliant on global connectivity.

Airports create direct jobs in operations, airlines, air traffic control, and security, while also generating indirect employment in sectors like retail, hospitality, and tourism. The International Air Transport Association (IATA) estimates that a 1 million increase in passenger traffic can generate up to 3,000 jobs (IATA, 2023). Green (2007) notes that airports often lead to the emergence of airport-related businesses, further stimulating local job markets. Yao et al. (2008) found a positive relationship between airport development and economic growth in China, alongside industrial structure, population density, and openness, although it showed a negative correlation with ground transportation. Similarly, Florida et al. (2012) concluded that airports significantly enhance economic output per capita in U.S. metropolitan areas, with the level of airport activity being a crucial factor.

2.2.2. Business Development and Investment Attraction, Tourism Development

Airports enhance regional competitiveness by improving accessibility, a crucial factor for businesses contemplating relocation or expansion. The presence of an international airport often influences multinational companies in choosing locations for regional headquarters or manufacturing facilities (Percoco, 2010). Airports can stimulate the growth of business parks, free trade zones, and industrial clusters nearby. For instance, Kasarda (2000) introduced the concept of the "aerotropolis," where the airport becomes a hub for economic activities such as logistics, retail, and commercial development. Ke et al. (2023) studied Brisbane Airport and found that the local urban context, rather than the airport's role as a transportation hub, is the primary factor attracting industries to airport areas. Airports also facilitate the development of knowledge-intensive industries and high-tech clusters by providing essential global connectivity that attracts talent and fosters innovation.

Goetz (1992) suggests that airports support research and development by connecting a wide array of suppliers, clients, and academic institutions, thus promoting technology transfer and innovation. Additionally, airports serve as gateways for domestic and international tourists, significantly impacting the tourism sector. Airport proximity is a key factor in a destination's attractiveness, with improved air connectivity linked to increased tourism revenues that support regional economies, especially those reliant on tourism (Forsyth et al., 2016). Studies show that airport infrastructure improvements lead to higher tourist arrivals and the development of supporting amenities like hotels and restaurants. Graham (2014) establishes a direct correlation between airport expansion and growth in local tourism-related businesses, while Tsiotas et al.

(2019) highlight the contribution of small and regional airports to tourism and regional development in Greece using a complex multiplier index.

2.2.3. Improved Accessibility and Social Integration

Airports enhance social development by improving access to remote regions, reducing geographic isolation, and facilitating social integration. This is particularly crucial in developing areas where air transport serves as the primary long-distance travel mode. The European Commission (2016) emphasizes that regional airports increase accessibility to less populated areas, making them attractive for economic development and reducing regional disparities. Xin et al. (2021) examined the spatial spillover effects of China's major hub airports on economic development and found that air passenger traffic, cargo shipments, and flight frequency create more substantial and positive spillover effects for neighboring cities connected to the airport network than the direct effects of the airport itself.

While existing literature mainly focuses on the impact of large hub airports in metropolitan areas, limited research explores their influence on peripheral or remote regions. Although metropolitan areas benefit from direct advantages like job creation and tourism, the extent of these benefits in less urbanized areas remains unclear. In rural regions, the effects of large hub airports may be less pronounced due to lower population density and limited infrastructure. Furthermore, the spillover benefits of a major airport may diminish with distance from urban centers. This gap in research underscores the need for further investigation into how large hub airports affect economic development in peripheral areas, where growth dynamics may differ from those in metropolitan contexts.

2.2.4. Reverse Causality: Airport Activity and Regional Development

The relationship between airports and regional development involves reverse causality, where airports stimulate economic growth while economically strong regions attract airport investments. Blonigen and Cristea (2015) argue that regions with higher economic activity and population growth draw more flights and larger airports, indicating a bidirectional relationship. Button and Taylor (2000) find that economically active regions tend to have larger airports, suggesting that regional factors can drive airport growth. Percoco (2010) highlights that prosperous regions often receive more investment in airport infrastructure, implying economic growth can cause airport expansion.

Similarly, Mukkala and Tervo (2013) demonstrate that air traffic and economic development are interconnected, with economically successful regions attracting increased air connectivity. Bai and Wu (2022) examine the relationship between airport development and regional economic growth in Jiangsu, China in 2008-2018 and found a bi-directional Granger causality between airport throughput and GDP, with economic growth having a stronger impact on airport development than vice versa. Both variables exhibited significant growth, with a sustainable and increasing interdependence over time. These studies underline the need to consider reverse causality when evaluating the effects of airports on regional development, as the relationship is often mutually reinforcing.

2.2.5. Denver International Airport

Denver International Airport (DIA) is one of the largest and busiest airports in the United States, serving as a major gateway to the Rocky Mountain region and beyond. Opened in 1995, DIA has since become a crucial transportation hub, facilitating over 69 million passengers annually (as of 2019) and providing a wide range of domestic and international flights. Its strategic location and extensive flight network connect Denver to key markets across the globe, contributing significantly to the local and regional economy.

The Denver metropolitan area, known as the Denver-Aurora-Lakewood Metropolitan Statistical Area (MSA), is a vibrant economic center characterized by a diverse economy that includes sectors such as technology, healthcare, finance, and tourism. This region has experienced rapid population growth and urban development, driven in part by its appeal as a desirable place to live and work. As a result, the metropolitan area has seen increased demand for air travel and airport services (Bureau of Economic Analysis, 2022).

Beyond the metropolitan region, DIA serves a broad array of surrounding areas in Colorado, including rural and remote regions. These areas may not share the same level of economic activity or population density as the urban center but are nonetheless impacted by the airport's operations. Communities located in peripheral regions, such as Colorado Springs, Fort Collins, and smaller towns in the surrounding counties in the Front Range region, rely on DIA for accessibility, transportation, and economic opportunities. The peripheral areas served by Denver International Airport, including rural communities and smaller cities, have different economic structures and sectors such as agriculture, tourism and outdoor recreation, mining and natural resources, small business and retail (Bureau of Economic Analysis, 2022).

Since this paper focuses on analyzing Denver International Airport, it is essential to consider the concept of reverse causality in the relationship between airports and regional development. The above-mentioned studies emphasize that this relationship is often mutually reinforcing. While DIA undoubtedly serves as a catalyst for economic growth in the Denver metropolitan area through increased connectivity, job creation, and tourism, it is also true that the region's economic vitality and population growth contributes to the airport's expansion and the demand for air services. This dynamic illustrates the complexity of the airport-development relationship in the context of DIA, highlighting the need for a detailed understanding of how economic factors can influence airport growth and vice versa.

Denver's expansive airport infrastructure has contributed to the region's economic growth, attracting businesses, enhancing tourism, and generating employment opportunities. Inbound and outbound air cargo traffic is closely related to international and domestic trade flows, serving as a critical channel for high-value and time-sensitive imports and exports. In Colorado, exports accounted for approximately 2% of state GDP in 2024, underscoring the importance of efficient transport infrastructure in maintaining global market access (Office of the U.S. Trade Representative, 2025). Denver International Airport (DEN), as a major cargo hub, plays a strategic role in supporting this trade activity—particularly for sectors like electronics, aerospace, and pharmaceuticals, which rely heavily on-air freight (Metro Denver EDC, 2023). While much of Colorado’s bulk trade still moves via trucks and trains due to cost considerations, air cargo provides a complementary mode for goods where speed and reliability are prioritized. These dynamics highlight the broader economic value of airport infrastructure not only for passenger movement but also for trade-related logistics.

But there are less studies that explored DIA’s effect on more rural and less densely populated areas. By investigating this relationship, we aim to fill a critical gap in the literature by providing insights into how a large hub airport like DIA impacts economic development in its surrounding peripheral regions of Colorado.

Given this context, our focus on exploring DIA is particularly relevant for several reasons. Firstly, Denver serves as a significant metropolitan hub with a vibrant economy, making it an ideal case study for examining the reverse causality between a major airport and urban economic development. Additionally, it is essential to extend this analysis beyond the metropolitan boundaries to assess how DIA influences the surrounding peripheral and remote areas of Colorado. By examining the relationship between Denver International Airport and the broader Colorado region, including both metropolitan and peripheral areas, we can gain valuable insights into how a major hub airport influences regional development dynamics. This context prepares a necessary need for our analysis of DIA's effects on economic development in both urban and regional settings. Also, this analysis is presented as a case study, which allows for a more detailed examination of the specific economic and regional dynamics associated with Denver International Airport. Focusing on a single, well-documented example enables a deeper exploration than would be feasible using broader, less-refined data across multiple airports with varying characteristics.

2.3. Data and Methods

The empirical analysis is conducted using monthly data related to airport transportation in Denver International Airport, and data related to economic growth in Denver Metropolitan Statistical Area (MSA) and in Colorado for the period of January 2000 to December 2019. In total, we have 240 monthly observations. By using these variables, it can be analyzed if there is a casual relationship between air transportation and local economic development in the Denver MSA region and in the whole Colorado region. In the analysis the data (in levels) is transformed into logs, because two different regions are analyzed (Denver MSA versus Colorado region).

2.3.1. Air Transportation and Economic Development Data

Data related to air transportation is downloaded from the official website of Denver International Airport (www.FlyDenver.com). Variables are:

- *Total Flights (number)* - An aircraft operation is either a landing or takeoff or contact with the control tower (as reported by the Federal Aviation Administration).
- *Total International Flights Passengers (number)* - International revenue and non-revenue passengers (as reported by the individual airline).
- *Total Domestic Flights Passengers (number)* - Domestic revenue and non-revenue passengers (as reported by the individual airline).
- *Total Cargo (pounds)* – Total inbound and outbound cargo transported to/from DIA.
- *Difference of Inbound-Outbound Cargo (pounds)* – Absolute net difference between inbound and outbound cargo transported to/from DIA.

As indicators for economic development in the Denver MSA region, two types of variables are used: employment and business entities. Variables are:

- *Total Non-Farm Employment in Denver MSA (thousands of people)* - All non-farm employees in Denver MSA region. Source: US Bureau of Labor Statistics (BLS.gov).
- *Total Non-Farm Employment in Colorado (thousands of people)* - All non-farm employees in the whole Colorado region. Source: US Bureau of Labor Statistics (BLS.gov).
- *Total Registered Business Entities in Denver MSA (number)* - All registered business entities in Denver MSA region. Source: State of Colorado (Colorado.gov).

2.3.2. Causality Analysis Framework

The Granger-causality econometric method is used to assess the cointegration and causality between air transportation and regional economic development. Cointegration occurs when two or more nonstationary variables share a stable long-term relationship, with deviations from this equilibrium adjusting over time. Cointegration tests examine non-stationary time series, which are processes where the means and variances change over time. Essentially, these tests enable the estimation of long-run parameters or equilibrium in systems where variables have unit roots. However, if no cointegration is detected, it does not rule out a relationship but suggests that one may not be present (Rao, 2007).

The idea of the Granger-causality test is to check the correlation between current values of one variable and past values of other variables (Granger, 1969). In other words, the past can impact the present and future, but the future cannot impact the present and past. For example, in the case of unidirectional causality, if lagged values of X and Y can predict X, then it is said that Y will cause X. If lagged values of X and Y can predict Y, then it is said that X will cause Y. Also, there might be the case of bi-directional causality if lagged values of both variables affect each other.

In the context of airport activity and regional economic development, causality can be theoretically caused by either one or both directions. There are four possible outcomes of the causality check regarding airport activity and regional development variables:

1. Air transportation causes regional economic development in both Denver MSA and Colorado regions.
2. Regional economic development in Denver MSA and Colorado regions causes development of air transportation in Denver International Airport.
3. Air transportation and regional economic development are causing bi-directional causality.

4. There is no causal relationship between air transportation and regional economic development.

To conduct a valid Granger-causality check, a four-step procedure is performed. First, stationarity tests are conducted on all air transportation and regional economic development variables. Performing the Granger-causality test requires that all variables must be stationary: mean and variance of variables remain constant/not constant over time. Otherwise, the results of the test might be spurious (Granger and Newbold, 1974). Then if data is not stationary, it is recommended to transform it by first differencing to make series stationary and it is said to be integrated of order one. The purpose of using the first-differencing method to stationarize time series data for air transportation and regional economic development variables is to ensure that sample statistics like means, variances, and correlations are meaningful. If the mean and variance of these series are not well-defined, then the correlations with other variables will also be unreliable.

Most of the air transportation variables are expected to demonstrate non-stationarity, while regional development variable is expected to be more stationarity. The reason for that might be because the air traffic variables are prone to more frequent changes with time due to frequent changes in air destinations, traffic, business needs and seasonality, while emerging businesses variable, which is a proxy for regional development, is expected to be more or less stable over time.

If after checking for stationarity it is revealed that variables are of mixed order then further cointegration of them is inappropriate, as this method requires all the variables to be integrated of order I(1). An autoregressive distributed lag (ARDL) model is used in the case of non-stationary time series as well as for times series with mixed order of integration.

For stationary variables, in contrast, air transportation and regional economic variables are tested with a cointegration test to reveal what kind of model would be applied. If the cointegration test reveals that series are not cointegrated, then Vector Autoregressive Model (VAR) would be applied. If cointegration test reveals that series are cointegrated (have a long-run relationship), then Vector Error Correction Model (VECM) would be applied, which would estimate both short-run and long-run causality parameters between the series (if any) (Toda and Phillips, 1993).

A diagram summarizing these procedures is presented in Figure-2.1.

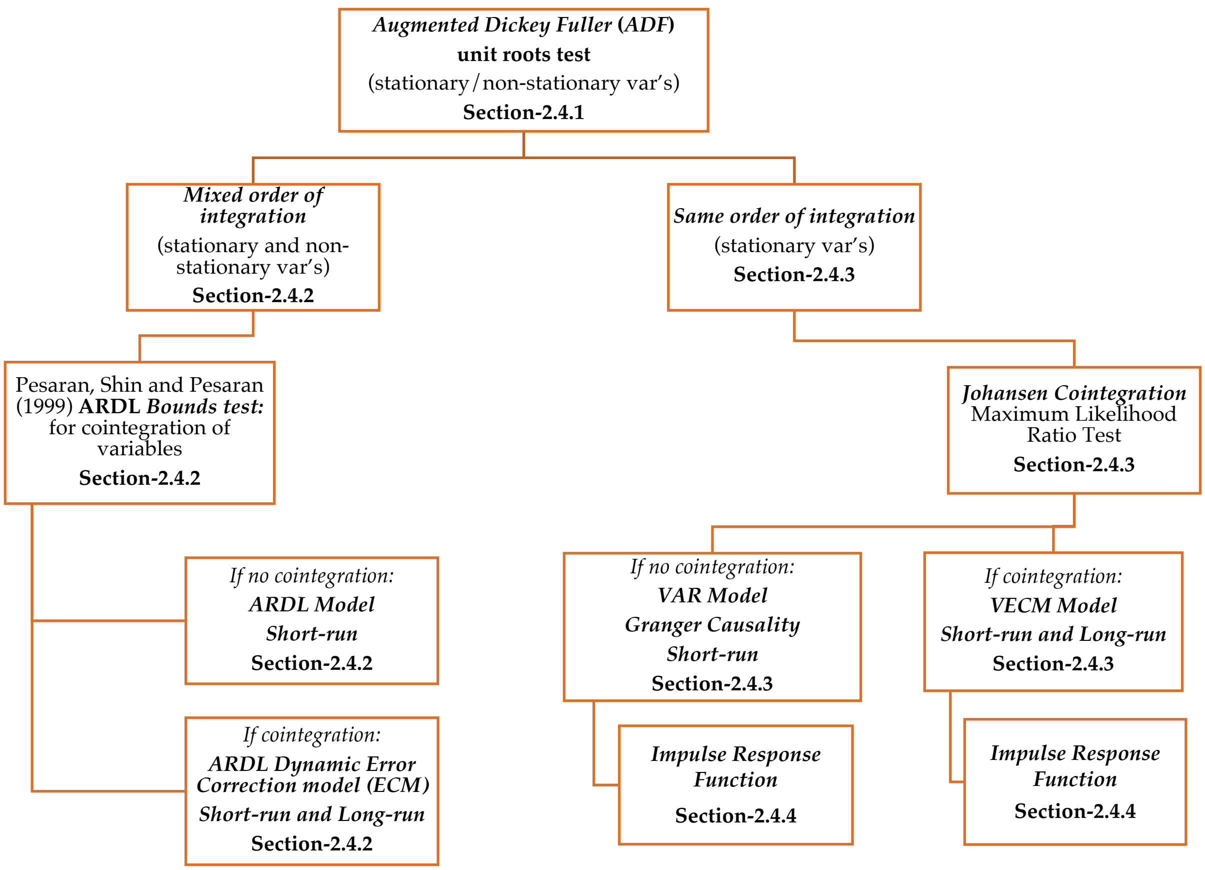


Figure-2.1. Granger-causality Framework

Source: Authors' elaboration.

2.3.2.1. Stationarity and the Order of Integration

The first step is to use the Augmented Dickey Fuller (ADF) test for detecting unit roots and if variables are stationary or non-stationary in the proposed specifications (Dickey & Fuller, 1981). The proposed specifications are:

$$\Delta\, Total\ Flights_t = \beta_0 + \beta_1 Total\ Flights_{t-1} + \beta_2 t + \sum_{i=1}^{p} \beta_i\, Total\ Flights_{t-i} + \varepsilon_t \tag{2.1}$$

$$\Delta\, Total\ Intl.Passengers_t = \delta_0 + \delta_1 Total\ Intl.Passengers_{t-1} + \delta_2 t + \\ + \sum_{i=1}^{p} \delta_i\, Total\ Intl.Passengers_{t-i} + \varepsilon_t \tag{2.2}$$

$$\Delta\, Total\ Domst.Passengers_t = \gamma_0 + \gamma_1 Total\ Domst.Passengers_{t-1} + \gamma_2 t + \\ \sum_{i=1}^{p} \gamma_i\, Total\ Domst.Passengers_{t-i} + \varepsilon_t \tag{2.3}$$

$$\Delta\, Total\ Cargo_t = \alpha_0 + \alpha_1 Total\ Cargo_{t-1} + \alpha_2 t + \sum_{i=1}^{p} \alpha_i\, Total\ Cargo_{t-i} + \varepsilon_t \tag{2.4}$$

$$\Delta\, Net\ Cargo_t = \lambda_0 + \lambda_1 Net\ Cargo_{t-1} + \lambda_2 t + \sum_{i=1}^{p} \lambda_i\, Net\ Cargo_{t-i} + \varepsilon_t \tag{2.5}$$

$$\Delta\, Total\ Employment\_Denver_t = \mu_0 + \mu_1 Total\ Employment\_Denver_{t-1} + \mu_2 t + \\ \sum_{i=1}^{p} \mu_i\, Total\ Employment\_Denver_{t-i} + \varepsilon_t \tag{2.6}$$

$$\Delta\, Total\ Employment\_Colorado_t = \eta_0 + \eta_1 Total\ Employment\_Colorado_{t-1} + \eta_2 t + \\ \sum_{i=1}^{p} \eta_i\, Total\ Employment\_Colorado_{t-i} + \varepsilon_t \tag{2.7}$$

The null (and alternative) hypothesis that tests the existence of unit root in the variables *Total Flights, Total International Flights Passengers, Total Domestic Flight Passengers, Total*

Cargo, Net Cargo, Total Employment in Denver MSA, Total Employment in Colorado region can be expressed as follows:

H_0: time series is not stationary	H_1: time series is stationary
$H_0: \beta_1 = 0$	$H_1: \beta_1 \neq 0$
$H_0: \delta_1 = 0$	$H_1: \delta_1 \neq 0$
$H_0: \gamma_1 = 0$	$H_1: \gamma_1 \neq 0$
$H_0: \alpha_1 = 0$	$H_1: \alpha_1 \neq 0$
$H_0: \lambda_1 = 0$	$H_1: \lambda_1 \neq 0$
$H_0: \mu_1 = 0$	$H_1: \mu_1 \neq 0.$
$H_0: \eta_1 = 0$	$H_1: \eta_1 \neq 0.$

Depending on the nature of the order of integration as determined by the ADF test, cointegration testing is then used to determine the appropriate model specifications.

2.3.2.2. Autoregressive Distributed Lag (ARDL) and Dynamic Error Corrections Models (ECM)

The ARDL cointegration approach was developed by Pesaran and Shin (1999) and Pesaran et al. (2001). An autoregressive distributed lag (ARDL) model is based on ordinary least square (OLS) model and is useful in the cases of both non-stationary time series and times series with mixed order of integration.

The ARDL bounds test is based on the joint F-statistic which its asymptotic distribution is non-standard under the null hypothesis of no cointegration. First, equations will be estimated using ordinary least squares (OLS). Then estimated equations will be tested for the existence of a long-run relationship among the variables by conducting an F-test for the joint significance of the coefficients of the lagged levels of the variables.

A dynamic error correction model (ECM) can be derived from ARDL through a simple linear transformation and both short-run and long-run dynamics will be integrated into a long-run equilibrium. ARDL error correction version modeling approach for airport transportation and regional development for cargo variables for Denver, for example, can be expressed as:

$$Net\ Cargo_t = \alpha + \beta Employment\_Denver_t + e_t \quad (2.8)$$

$$\Delta Net\ Cargo_t = \alpha_0 + \sum_{i=1}^{p} \beta_i\, \Delta Net\ Cargo_{t-i} + \sum_{i=1}^{p} \delta_i\, \Delta Employment\_Denver_{t-i} + \lambda_1 Net\ Cargo_{t-1} + \lambda_2 Employment\_Denver_{t-1} + u_t \quad (2.9)$$

For Colorado:

$$Net\ Cargo_t = \alpha + \beta Employment\_Colorado_t + e_t \quad (2.10)$$

$$\Delta Net\ Cargo_t = \alpha_0 + \sum_{i=1}^{p} \beta_i\, \Delta Net\ Cargo_{t-i} + \sum_{i=1}^{p} \delta_i\, \Delta Employment\_Colorado_{t-i} + \lambda_1 Net\ Cargo_{t-1} + \lambda_2 Employment\_Colorado_{t-1} + u_t \quad (2.11)$$

where: β, δ - represents short run dynamics of the model. λ_s - represents long run relationship. $H_0: \lambda_1 + \lambda_2 = 0$, which means non-existence of long run relationship.

2.3.2.3. Vector Autoregressive (VAR) and Vector Error Correction (VECM) Models

The Vector Autoregressive (VAR) model can be used to reveal the reverse causality among the airport transportation variables and regional development variables using their own past values. In VAR model, all variables are considered endogenous. For Denver, the specifications take the forms:

$$Total\ Employment\_Denver_t = \delta_1 + \theta_{11} Total\ Employment\_Denver_{t-1} + \theta_{12} Total\ Cargo_{t-1} + \varepsilon_{1t} \quad (2.12)$$

$$Total\ Cargo_t = \delta_2 + \theta_{21} Total\ Employment\_Denver_{t-1} + \theta_{22} Total\ Cargo_{t-1} + \varepsilon_{2t} \quad (2.13)$$

Analogously for Colorado:

$$Total\ Employment\_Colorado_t = \lambda_1 + \phi_{11} Total\ Employment\_Colorado_{t-1} + \phi_{12} Total\ Cargo_{t-1} + \epsilon_{1t} \quad (2.14)$$

$$Total\ Cargo_t = \lambda_2 + \phi_{21} Total\ Employment\_Colorado_{t-1} + \phi_{22} Total\ Cargo_{t-1} + \epsilon_{2t} \quad (2.15)$$

where, *Total Employment_Denver* or *Total Employment_Colorado* and *Total Cargo* are three variables with only one lag, ε_{1t} are ε_{2t}, ϵ_{1t} are ϵ_{2t} - are uncorrelated white noise disturbances or error terms. The optimal number of lags can be selected by using Akaike Information Criterion (AIC), Schwartz Bayesian Criterion (SBC), and Hannan Quinn criterion (HQC).

Engel and Granger (1987) demonstrated that, if two variables are integrated of the same order and cointegrated, then there might exist a causal long-run relationship between variables in at least one direction. Cointegration test reveals if there exists a long-run equilibrium relationship between variables. If the test shows that there is no cointegration between variables, then it means that variables do not have long-run equilibrium relationship between each other (Dickey, Jansen, & Fuller, 1991).

The Johansen cointegration maximum likelihood ratio test (1988) is the most commonly used test, and it is based on the null hypothesis of no-cointegration against an alternative hypothesis of cointegration. The results of the cointegration test would suggest which model should be used for further Granger-causality testing. The positive cointegration test reveals that there is a long-run equilibrium relationship between variables, but it does not indicate the direction of the causal relationship.

If cointegration is not detected, then Vector Autoregressive (VAR) model should be used, with further Wald test application for Granger-causality (Engel and Granger, 1987). If cointegration is detected, then Vector Error Correction Model (VECM) should be used (VAR is unspecified in the presence of cointegration), which would test both short-run and long-run causality relationship between the series (Toda and Phillips, 1993).

According to Engel and Granger (1987), if two series are cointegrated, then VECM model for air transportation variables in Denver International Airport (proxied by *Domestic Passengers*) and regional economic development variables (proxied by *Total Business Entities registered in Denver MSA region*) series can be written as follows:

$$\Delta Registered\ Business\_Denver_t = \alpha_{1j} + \eta_1 ECT_{t-1} + \sum_{k=1}^{p} \beta_k\, \Delta Registered\ Business\_Denver_{t-k} + \sum_{k=1}^{p} \lambda_k \Delta Domestic\ Passengers_{t-k} + \varepsilon_{1t} \quad (2.16)$$

$$\Delta Domestic\ Passengers_t = \alpha_{1j} + \eta_2 ECT_{t-1} + \sum_{k=1}^{p} \beta_k\, \Delta Domestic\ Passengers_{t-k} + \sum_{k=1}^{p} \lambda_k \Delta Registered\ Business\_Denver_{t-k} + \varepsilon_{2t} \qquad (2.17)$$

where, $\Delta Registered\ Business_{Denver}$ and $\Delta Domestic\ Passengers$ denote first differencing in regional development and air transportation variables that capture their short-run disturbances over periods $t = 1, 2, \ldots T$; ε_{1t} and ε_{2t} are the serially uncorrelated error terms and ECT_{t-1} is the error correction term (ECT), resulted from the long-run cointegration relationship, and assesses the extent of the past disequilibrium.

The coefficient, η of the ECT_{t-1}, represents the deviation of the dependent variables from the long-run equilibrium (the estimated residual from the cointegration regression), and p is the number of lags. A significant coefficient implies that past equilibrium errors play a role in determining the current outcomes and there exists long-run equilibrium relationship affecting the dependent variable. The error correction term (speed of adjustment) gives a magnitude of speed with which the cointegrating variables return to the long-run traverse following any short-run distortion. The *ECT* coefficient is expected to be a negative number, between 0 and 1 as an absolute number. If coefficient is positive – this means that there exist some errors with estimation of causality.

The short run dynamics are captured through the individual coefficients of the difference terms. To obtain the estimated residuals *(ECT)* a long-run model was run first, then estimated residual were later used to estimate the Granger causality model with a dynamic error correction (ECT_{t-1}). Equation (2.16) shows that if the information contained in the past values of $\Delta Total\ Domestic\ Flights\ Passengers\ (DIA)$ is significantly able to explain $\Delta Total\ Registered\ Businesses\ in\ Denver\ MSA$ with the past values of

$\Delta Total\ Registered\ Businesses\ in\ Denver\ MSA$, then we can say that $\Delta Total\ Domestic\ Flights\ Passengers\ (DIA)$ Granger causes $\Delta Total\ Registered\ Businesses\ in\ Denver\ MSA$. In other words, a change in the level of *Total Domestic Flights Passengers (DIA)* can lead to a change in the level of *Total Registered Businesses in Denver MSA*.

Accordingly, Equation (2.17) shows that if the information contained in the past values of $\Delta Total\ Registered\ Businesses\ in\ Denver\ MSA$ is significantly able to explain $\Delta Total\ Domestic\ Flights\ Passengers\ (DIA)$ with the past values of $\Delta Total\ Domestic\ Flights\ Passengers\ (DIA)$, then we can say that $\Delta Total\ Registered\ Businesses\ in\ Denver\ MSA$ Granger-causes $\Delta Total\ Domestic\ Flights\ Passengers\ (DIA)$.

So, we can say that a change in the level of *Total Registered Businesses in Denver MSA* can lead to a change in the level of *Total Domestic Flights Passengers (DIA)*. Also, in each equation, the previous period's disequilibrium in level (ECT_{t-1}) also causes the change in the endogenous variable *Total Registered Businesses in Denver MSA*. Because of this error-correction representation it is possible to distinguish short-run and long-run causality separately.

2.3.2.4. Impulse Response Functions

The relationships between variables in a VAR model are not immediately clear from the parameter matrices alone. As a result, impulse response functions (IRFs) are commonly used to interpret VAR models. According to Lütkepohl (2010), a general VAR model can be expressed more compactly as $A(L)y_t = u_t$, where L is the lag operator, defined such that $Ly_t = y_{t-1}$ and $A(L) = I_K - A_1 L - A_p L^p$ is a matrix polynomial in the lag operator. If the polynomial $\det A(z)$ has all its roots outside the complex unit circle, the process is stationary and has a Wold moving average (MA) representation:

$$y_t = A(L)^{-1}u_t = u_t + \sum_{i=1}^{\infty} \Phi_i u_{t-i} \quad (2.18)$$

In this context, impulse response analysis involves a counterfactual approach, where the effect of a shock to one variable is traced through the system. This is done by setting a single component of u_t to one, while keeping all other components at zero, and then observing how the responses of y_t evolve over time. The resulting impulse responses correspond to the elements of the $\Phi_i$matrices. Since u_t represents the one-step-ahead forecast errors, these functions are often referred to as *forecast error impulse responses* (for example, see Lütkepohl, 2005, Sec. 2.3.2).

Such a counterfactual experiment may not accurately reflect the actual responses of an economic system because the components of u_t can be correlated at the same time, meaning Σ_u may not be a diagonal matrix. In this case, forecast error impulses may not be realistic, as a shock to one variable is likely to trigger a response in other variables simultaneously, which means they cannot be treated in isolation.[1] To address this issue, *orthogonalized impulse responses* are commonly used. These are derived by selecting a matrix B such that $BB' = \Sigma_u$ or equivalently, $B^{-1}\Sigma_u B'^{-1}$ is a diagonal matrix and defining $\varepsilon_t = B^{-1}u_t$. Substituting in (2.18) gives

$$y_t = B\varepsilon_t + \sum_{i=1}^{\infty} \Theta_i \varepsilon_{t-i} \quad (2.19)$$

where, $\Theta_i = \Phi_i B,\ i = 1, 2, \dots$. The ε_t's are now contemporaneously uncorrelated (orthogonal), with a diagonal or unit covariance matrix. This approach ensures that the shocks to, ε_t provide a more

[1] Non-orthogonalized impulse responses are sensitive to the ordering of variables in VAR. This involves a "researcher degrees of freedom" problem, because the ordering of variables in the estimator is arbitrary (Tan et al, 2025).

accurate (not sensitive to the ordering of variables in the VAR) and realistic picture of how the system responds to external disturbances.

2.4. Empirical Results

The summary statistics of the data used in the analysis are presented in Table-2.1.

Table-2.1. Summary Statistics

Variable	Obs.	Mean	Std.Dev	Min	Max
Air Transportation Variables, Denver International Airport					
Total Flights, number	239	48,519	4,694	3,4155	59,716
Total International Flights Passengers, number	239	118,353	45,064	44,696	218,555
Total Domestic Flights Passengers, number	239	4,047,238	821,691	1,928,422	6,417,061
Total Cargo, pounds	239	52,200,000	11,200,000	35,700,000	95,400,000
Difference of Inbound-Outbound Cargo (absolute value), pounds	239	5,763,896	2,045,627	999,826	13,800,000
Employment Variable					
Total Non-Farm Employment in Denver MSA, thousands people	240	1,282.87	117.95	1,153.9	1,554.3
Total Non-Farm Employment in Colorado region, thousands people	240	2,366.79	192.11	2,145.6	2,811.5
Business Entities					
Total Registered Businesses in Denver MSA per month, number	240	96.45	101.15	16	1,200

Source: Authors' calculations based on data from above-mentioned sources.

Seasonal adjustments were required to remove any seasonal effects that may be present in the dataset. Figures A1-A7 in *the Appendix-A* show the raw and seasonally adjusted data of airport related and economic variables. In the analysis, we use seasonally adjusted data. The airport variables are hugely affected by seasonality, while business variables like employment both in the Colorado region and Denver MSA, and the number of business in Denver MSA does not depend on seasonality so heavily.

2.4.1. Stationarity Unit Root Test Results

To proceed with Granger causality, it is important to fulfill the necessary pre-test condition for cointegration test: each of the series should be integrated of the *same* order (more than zero), or that both series should contain a deterministic trend (Granger, 1988). To verify this condition, the Augmented Dickey Fuller Panel (ADF) Unit Root Test as described above is performed for each of the Airport Transportation and Regional Development variables in the dataset (Table-2.2).

Table-2.2. Augmented Dickey Fuller Panel Unit Root Test Statistics

Variable	Log-transformed form		First-difference		Stationarity
	Statistics[1]	Probability	Statistics	Probability	
Total Non-Farm Employment in Denver MSA, thousands people	-1.02 (4)	0.94	-5.45 (4)	0.00	Stationary of degree (1)
Total Non-Farm Employment in Colorado, thousands people	-0.733 (4)	0.97	-5.09 (2)	0.00	Stationary of degree (1)
Total Registered Businesses in Denver MSA per month, number	-9.48 (4)	0.00	-16.35 (4)	0.00	Stationary of degree (0)
Total Flights, number	-2.46 (4)	0.34	-13.56 (4)	0.00	Stationary of degree (1)
Total International Flights Passengers, number	-1.70 (4)	0.74	-12.84 (4)	0.00	Stationary of degree (1)
Total Domestic Flights Passengers, number	-4.58 (4)	0.001	-14.16 (4)	0.00	Stationary of degree (0)
Total Cargo, pounds	-1.92 (4)	0.64	-20.01 (4)	0.00	Stationary of degree (1)
Difference of Inbound-Outbound Cargo (absolute value), pounds	-5.43 (4)	0.00	-17.44 (4)	0.00	Stationary of degree (0)

[1]Figures in parentheses represent lag length based on the Akaike Information Criterion.

Source: Authors' calculations.

2.4.2. ARDL and ECM Model Results

The ADF stationarity test revealed that *not all* variables are stationary (integrated of I(0) order). We cannot proceed with the Johansen cointegration test with all of the pairs of air transportation and regional development variables, because of mixed order. For Johansen cointegration, the test can be performed only with those pair of variables who demonstrated the same order of integration: either I(0) or I(1).

For remaining pairs of variables with mixed order we should proceed with an Autoregressive Distributed Lag (ARDL) model. The ARDL model can capture both long-run and short-run relations of the cointegrated variables of mixed order. To test the long run cointegration of variables in ARDL model two sets of critical values for a given significance level can be determined (Pesaran et al., 2001). The first level is calculated on the assumption that all variables included in the ARDL model are integrated of order zero, while the second one is calculated on the assumption that the variables are integrated of order one. The null hypothesis of no cointegration is rejected when the value of the test statistic exceeds the upper critical bounds value, while it is accepted if the F-statistic is lower than the lower bounds value. Otherwise, the cointegration test is inconclusive. The results of ARDL modeling approach are presented below in Tables 2.3-2.11.

Testing Pair-1: ***Total Non-Farm Employment (Denver MSA and Colorado region) vs. Total Domestic Flights Passengers (DIA)***

The first step is to determine if these two variables have long-run cointegration. To determine that the ARDL Bound test (4, 4) is estimated and results are presented in Table-2.3.

Table-2.3. ARDL Bound Test Results

Denver MSA Region

Variables	Status of var's	AIC lag	F-statistic	Decision
Ln_Total Non-Farm Employment in Denver MSA	Dependent	4	1.873	Accept H_0. No long-run cointegration.

Ln_Total Domestic Flights Passengers	Independent			Proceed with short-run ARDL model.
Ln_Total Domestic Flights Passengers	Dependent	4	1.492	Accept H_0. No long-run cointegration. Proceed with short-run ARDL model.
Ln_Total Non-Farm Employment in Denver MSA	Independent			

Note: Lower and Upper-bound critical values are taken from Pesaran et al (2001).

Colorado Region

Variables	Status of var's	AIC lag	F-statistic	Decision
Ln_Total Non-Farm Employment	Dependent	4	1.032	Accept H_0. No long-run cointegration. Proceed with short-run ARDL model.
Ln_Total Domestic Flights Passengers	Independent			
Ln_Total Domestic Flights Passengers	Dependent	4	1.829	Accept H_0. No long-run cointegration. Proceed with short-run ARDL model.
Ln_Total Non-Farm Employment	Independent			

Note: Lower and Upper-bound critical values are taken from Pesaran et al (2001).

Source: Authors' calculations.

For the case, where *Total Non-farm Employment* is defined as a dependent variable, the ARDL bound test result showed that there is not a long-run cointegration between *Total Non-farm Employment (both* in Denver MSA and Colorado region) and *Total Domestic Flights Passengers*, then the short-run ARDL modeling estimation is performed (Table-2.4). The ARDL (4, 4) model for Denver MSA and ARDL (4, 3) model for Colorado region are selected based on Akaike Information Criterion.

For the reverse case, where *Total Domestic Flights Passengers* is defined as a dependent variable, the above ARDL bound test showed that there is not a long-run cointegration between *Total Domestic Flights Passengers* and *Total Non-farm Employment (both* in Denver MSA and Colorado region*)*, the short-run ARDL modeling estimation is performed (Table-2.4). The ARDL (4, 4) model for Denver MSA and ARDL (3, 0) model for Colorado region are selected based on Akaike Information Criterion.

Table-2.4. ARDL Model Results

Denver MSA Region

Short-Run ARDL Model Results	(1)
Variables	***Ln_Total Non-Farm Employment in Denver MSA***
L. Ln_Total Non-Farm Employment in Denver MSA	1.193***
	(0.0633)
L2. Ln_Total Non-Farm Employment in Denver MSA	0.0639
	(0.102)
L3. Ln_Total Non-Farm Employment in Denver MSA	0.0479
	(0.103)
L.4 Ln_Total Non-Farm Employment in Denver MSA	-0.307***
	(0.0634)
Ln_Total Domestic Flights Passengers	0.00689**
	(0.00290)
L. Ln_Total Domestic Flights Passengers	0.00203
	(0.00348)
L2. Ln_Total Domestic Flights Passengers	-0.00365
	(0.00352)
L3. Ln_Total Domestic Flights Passengers	-0.00140
	(0.00347)
L4. Ln_Total Domestic Flights Passengers	-0.00166
	(0.00295)
Constant	-0.0149
	(0.0118)
Observations	236
R-squared	1.000

Short-Run ARDL Model Results	(2)
Variables	***Ln_Total Domestic Flights Passengers***
L. Ln_Total Domestic Flights Passengers	0.645***
	(0.0662)
L2. Ln_Total Domestic Flights Passengers	0.140*
	(0.0792)
L3. Ln_Total Domestic Flights Passengers	0.0985
	(0.0783)
L4. Ln_Total Domestic Flights Passengers	0.0707
	(0.0666)
Ln_Total Non-Farm Employment in Denver MSA	3.528**
	(1.486)
L. Ln_Total Non-Farm Employment in Denver MSA	-3.740
	(2.283)
L2. Ln_Total Non-Farm Employment in Denver MSA	1.703
	(2.299)
L3. Ln_Total Non-Farm Employment in Denver MSA	-5.177**
	(2.304)
L4. Ln_Total Non-Farm Employment in Denver MSA	3.768**
	(1.486)
Constant	0.110
	(0.269)
Observations	236
R-squared	0.951

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Colorado Region

Short-Run ARDL Model Results	(3)
Variables	***Ln_Total Non-Farm Employment in Colorado region***

Short-Run ARDL Model Results	(4)
Variables	***Ln_Total Domestic Flights Passengers***

L. Ln_Total Non-Farm Employment in Colorado region	1.230***
	(0.0637)
L2. Ln_Total Non-Farm Employment in Colorado region	0.0586
	(0.103)
L3. Ln_Total Non-Farm Employment in Colorado region	-0.0231
	(0.103)
L4. Ln_Total Non-Farm Employment in Colorado region	-0.268***
	(0.0635)
Ln_Total Domestic Flights Passengers	0.00328
	(0.00263)
L. Ln_Total Domestic Flights Passengers	0.00111
	(0.00313)
L2. Ln_Total Domestic Flights Passengers	0.00153
	(0.00313)
L3. Ln_Total Domestic Flights Passengers	-0.00438*
	(0.00263)
Constant	-0.00507
	(0.0118)
Observations	236
R-squared	1.000

L. Ln_Total Domestic Flights Passengers	0.665***
	(0.0651)
L2. Ln_Total Domestic Flights Passengers	0.157**
	(0.0778)
L3. Ln_Total Domestic Flights Passengers	0.126*
	(0.0655)
Ln_Total Non-Farm Employment in Colorado region	0.113*
	(0.0638)
Constant	-0.0930
	(0.274)
Observations	237
R-squared	0.949

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Source: Authors' calculations.

The short-run estimates of ARDL show that *Total Domestic Flights Passengers* variable is the determinant of *Total Non-Farm Employment* both in Denver MSA and in Colorado region in the short-run (in log form). According to the test results, a one percent point change in the *Total Domestic Flights Passengers* brings a change of about 0.006 percent point *increase* in *Total Non-Farm Employment* in Denver MSA and about 0.004 percent point *decrease* in *Total Non-Farm Employment* in Colorado region *(Lag-3)*, on average, ceteris paribus. The short-run estimates of

ARDL also show that *Total Non-Farm Employment* variable in both Denver MSA and Colorado regions is the determinant of *Total Domestic Flights Passengers* in the short-run (in log form).

According to the test results for Denver MSA: in Lag-3, one percent point change in *Total Non-Farm Employment* in Denver brings a change of about 5.17 percent point *decrease* in *Total Domestic Flights Passengers* in DIA; in Lag-4, a one percent point change in *Total Non-Farm Employment* in Denver brings a change of about 3.76 percent point *increase* in *Total Domestic Flights Passengers* in DIA, on average, ceteris paribus.

According to the results for Colorado region: one percent point change in *Total Non-Farm Employment* in Colorado brings a change of about 0.113 percent point *increase* in *Total Domestic Flights Passengers* in DIA, on average, ceteris paribus. To estimate the validity of the estimated models several diagnostic tests are performed and results are presented in *Appendix-B*[2].

In general, analyzing the pair of *Total Non-Farm Employment* and *Total Domestic Flights Passengers* give the following result in Table-2.5: both variables demonstrate bi-directional short-run causality, with higher magnitude effect from *Total Employment* variable both in Denver and Colorado regions on *Domestic Passengers* variable at DIA.

Table-2.5. Pair-1 Results

Denver MSA Region		
Dependent variable	**Independent variable**	**Causality**
Total Non-Farm Employment in Denver MSA	*Total Domestic Flights Passengers*	Short-run, positive, significant, lower magnitude effect
Total Domestic Flights Passengers	*Total Non-Farm Employment in Denver MSA*	Short-run (lags-3,4), positive/negative, significant, higher magnitude effect
Colorado Region		
Dependent variable	**Independent variable**	**Causality**
Total Non-Farm Employment in Colorado region	*Total Domestic Flights Passengers*	Short-run, negative (Lag-3), significant, lower magnitude effect

[2] Appendix-B (Tables-B1-B4).

Total Domestic Flights Passengers	*Total Non-Farm Employment in Colorado region*	Short-run, positive, significant, higher magnitude effect

Source: Authors' elaborations.

Testing Pair-2: *Total Non-Farm Employment (Denver MSA and Colorado region) vs. Difference of Inbound-Outbound Cargo (DIA)*

The first step is to determine if these two variables have long-run cointegration. To determine that the ARDL Bound test is estimated and results are presented in Table-2.6.

Table-2.6. ARDL Bound Test Results

Denver MSA Region

Variables	Status of var's	AIC lag	F-statistic	Decision
Ln_Total Non-Farm Employment in Denver MSA	Dependent	4	0.267	Accept H_0. No long-run cointegration. Proceed with short-run ARDL model.
Ln_Net In-Out Cargo	Independent			
Ln_Net In-Out Cargo	Dependent	4	10.725	Reject H_0. There is a long-run cointegration. Proceed with long-run ARDL model.
Ln_Total Non-Farm Employment in Denver MSA	Independent			

Note: Lower and Upper-bound critical values are taken from Pesaran et al (2001).

Colorado Region

Variables	Status of var's	AIC lag	F-statistic	Decision
Ln_Total Non-Farm Employment in Colorado region	Dependent	4	0.652	Accept H_0. No long-run cointegration. Proceed with short-run ARDL model.
Ln_Net In-Out Cargo	Independent			
Ln_Net In-Out Cargo	Dependent	4	10.780	Reject H_0. There is a long-run cointegration. Proceed with long-run ARDL model.
Ln_Total Non-Farm Employment in Colorado region	Independent			

Note: Lower and Upper-bound critical values are taken from Pesaran et al (2001).

Source: Authors' calculations.

For the case, where *Total Non-farm Employment* is defined as a dependent variable, the ARDL bound test result showed that there is not a long-run cointegration between *Total Non-farm*

Employment (both in Denver MSA and Colorado region) and *Net In-Out Cargo in DIA*, then the short-run ARDL modeling estimation is performed (Table-2.7). The ARDL (4, 3) model for Denver MSA and ARDL (3, 4) model for Colorado region are selected based on Akaike Information Criterion.

For the reverse case, where *Net In-Out Cargo in DIA* is defined as a dependent variable, the above ARDL bound test showed that *there is* a long-run cointegration between *Net In-Out Cargo in DIA* and *Total Non-farm Employment (both* in Denver MSA and Colorado region*)*, so the long-run ARDL modeling estimation is performed (Table-2.7). The ARDL (3, 3) model for Denver MSA and ARDL (3, 0) model for Colorado region are selected based on Akaike Information Criterion.

Table-2.7. ARDL Model Results

Denver MSA Region

Short-Run ARDL Model Results	(5)
Variables	***Ln_Total Non-Farm Employment in Denver MSA***
L. Ln_Total Non-Farm Employment in Denver MSA	1.216***
	(0.0635)
L2. Ln_Total Non-Farm Employment in Denver MSA	0.0728
	(0.103)
L3. Ln_Total Non-Farm Employment in Denver MSA	0.00597
	(0.104)
L4. Ln_Total Non-Farm Employment in Denver MSA	-0.295***
	(0.0643)
Ln_Net In-Out Cargo	0.000818
	(0.000602)
L1. Ln_Net In-Out Cargo	0.000694
	(0.000581)
L2. Ln_Net In-Out Cargo	-0.00107*
	(0.000572)
L3. Ln_Net In-Out Cargo	-8.66e-05
	(0.000573)
Constant	-0.00572
	(0.0109)
Observations	236
R-squared	1.000

Long-Run ARDL Model Results	(6)
Variables	***Ln_Net In-Out Cargo***
ADJ *L1. Ln_Net In-Out Cargo*	-0.360***
	(0.0779)
LR *Ln_Total Non-Farm Employment in Denver MSA*	2.240***
	(0.466)
SR	
LD. Ln_Net In-Out Cargo	-0.456***
	(0.0831)
L2D. Ln_Net In-Out Cargo	-0.215***
	(0.0803)
L3D. Ln_Net In-Out Cargo	-0.0902
	(0.0626)
D. Ln_Total Non-Farm Employment in Denver MSA	8.649
	(7.259)
LD. Ln_Total Non-Farm Employment in Denver MSA	1.220
	(7.150)
L2D. Ln_Total Non-Farm Employment in Denver MSA	-17.11**
	(7.146)
L3D. Ln_Total Non-Farm Employment in Denver MSA	7.091
	(7.355)
Constant	-0.168
	(1.198)
Observations	236
R-squared	0.450

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Colorado Region

Short-Run ARDL Model Results	(7)
Variables	***Ln_Total Non-Farm Employment***
L. Ln_Total Non-Farm Employment	1.325***
	(0.0615)

Long-Run ARDL Model Results	(8)
Variables	***Ln_Net In-Out Cargo***
ADJ *L1. Ln_Net In-Out Cargo*	-0.363***
	(0.0785)

L2. Ln_Total Non-Farm Employment	0.0585
	(0.108)
L3. Ln_Total Non-Farm Employment	-0.382***
	(0.0609)
Ln_Net In-Out Cargo	0.000438
	(0.000554)
L1. Ln_Net In-Out Cargo	0.000339
	(0.000534)
L2. Ln_Net In-Out Cargo	-0.00103*
	(0.000541)
L3. Ln_Net In-Out Cargo	0.000255
	(0.000535)
L4. Ln_Net In-Out Cargo	0.000197
	(0.000529)
Constant	-0.0124
	(0.0122)
Observations	231
R-squared	1.000

LR Ln_Total Non-Farm Employment	2.450***
	(0.481)
SR	
LD. Ln_Net In-Out Cargo	-0.459***
	(0.0835)
L2D. Ln_Net In-Out Cargo	-0.231***
	(0.0806)
L3D. Ln_Net In-Out Cargo	-0.110*
	(0.0630)
Ln_Total Non-Farm Employment	0.890***
	(0.248)
Constant	-1.258
	(1.361)
Observations	231
R-squared	0.437

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Source: Authors' calculations.

The short-run estimates of ARDL show that *Difference of Inbound-Outbound Cargo (DIA)* variable affects *Total Non-Farm Employment* both in Denver MSA and Colorado region in the short-run (in log form). According to the test results, a one percent point change in *Difference of Inbound-Outbound Cargo (DIA)* brings a change of about 0.001 percent point *decrease* in *Total Non-Farm Employment* both in Denver MSA and in Colorado, on average, ceteris paribus *(Lag-2)*. The long-run estimates of ARDL show that *Total Non-Farm Employment* variable in both Denver MSA and Colorado regions effects *Difference of Inbound-Outbound Cargo (DIA)* in the long-run (in log form).

According to the test results for Denver MSA: *Long-run:* one percent point change in *Total Non-Farm Employment* in Denver associates with a change of about 2.24 percent point *increase* in *Difference of Inbound-Outbound Cargo* in DIA, on average, ceteris paribus. The adjustment term shows that errors in the previous period will be corrected in the current period, the coefficient is negative (-0.360) and statistically significant at the 1% level. So, 36% of the equilibrium deviation will be corrected in the current period. *Short-run in Lag-2:* one percent point change in *Total Non-Farm Employment* in Denver brings a change of about 17.11 percent point *decrease* in *Difference of Inbound-Outbound Cargo (DIA)*, on average, ceteris paribus.

According to the results for Colorado region: *Long-run:* one percent point change in *Total Non-Farm Employment* in Colorado associates with a change of about 2.45 percent point *increase* in *Difference of Inbound-Outbound Cargo* in DIA, on average, ceteris paribus. The adjustment term shows that errors in the previous period will be corrected in the current period, the coefficient is negative (-0.363) and statistically significant at 1% level. So, 36.3% of the equilibrium deviation will be corrected in the current period. *Short-run:* one percent point change in *Total Non-Farm Employment* in Colorado brings a change of about 0.89 percent point *increase* in *Difference of Inbound-Outbound Cargo (DIA)*, on average, ceteris paribus.

To estimate the validity of the estimated models several diagnostic tests are performed and results are presented in *Appendix-B*[3].

In general, analyzing the pair of *Total Non-Farm Employment* and *Difference of Inbound-Outbound Cargo (DIA)* reveals the following result in Table-2.8: both variables demonstrate bi-directional short-run/long-run causality, with higher magnitude effect from Total Employment variable both in Denver and Colorado regions on Net Cargo variable at DIA.

[3] Appendix-B (Tables-B1-B4).

Table-2.8. Pair-2 Results

Denver MSA Region		
Dependent variable	**Independent variable**	**Causality**
Total Non-Farm Employment in Denver MSA	*Difference of Inbound-Outbound Cargo (DIA)*	Short-run (lag-2), negative, significant, lower magnitude effect
Difference of Inbound-Outbound Cargo (DIA)	*Total Non-Farm Employment in Denver MSA*	Long-run, positive, significant Short-run (lag-2), negative, significant, higher magnitude effect
Colorado Region		
Dependent variable	**Independent variable**	**Causality**
Total Non-Farm Employment in Colorado region	*Difference of Inbound-Outbound Cargo (DIA)*	Short-run, negative (Lag-2), significant, lower magnitude effect
Difference of Inbound-Outbound Cargo (DIA)	*Total Non-Farm Employment in Colorado region*	Long-run, positive, significant Short-run, positive, significant, higher magnitude effect

Source: Authors' elaborations.

Testing Pair-3: *Total International Flights Passengers (DIA) vs. Total Registered Businesses in Denver MSA*

The first step is to determine if these two variables have long-run cointegration. To determine that the ARDL Bound test is estimated and results are presented in Table-2.9.

Table-2.9. ARDL Bound Test Result

Variables	Status of var's	AIC lag	F-statistic	Decision
Ln_Total Registered Businesses	Dependent	4	2.836	Accept H_0. No long-run cointegration. Proceed with short-run ARDL model.
Ln_Total International Flights Passengers	Independent			
Ln_Total International Flights Passengers	Dependent	4	1.068	Accept H_0. No long-run cointegration. Proceed with short-run ARDL model.
Ln_Total Registered Businesses	Independent			

Lower and Upper-bound critical values are taken from Pesaran et al (2001).

Source: Authors' calculations.

Since the ARDL bound test result showed that there is no long-run cointegration between *Total Registered Businesses in Denver MSA* and *Total International Flights Passengers (DIA)*, the short-run ARDL modeling estimation is performed (Table-2.10). The ARDL (3, 1) model was selected based on Akaike Information Criterion. For reverse case, since above ARDL bound test result showed that there is no long-run cointegration between *Total International Flights Passengers (DIA)* and *Total Registered Businesses in Denver MSA*, the short-run ARDL modeling estimation was performed (Table-2.10). The ARDL (3, 3) model was selected based on Akaike Information Criterion.

Table-2.10. ARDL Model Results

Short-Run ARDL Model Results	(9)
Variables	***Ln_Total Registered Businesses***
L. Ln_Total Registered Businesses	0.415***
	(0.0628)
L2. Ln_Total Registered Businesses	0.342***
	(0.0644)
L3. Ln_Total Registered Businesses	0.151***
	(0.0569)
Ln_Total International Flights Passengers	0.837***
	(0.236)
L. Ln_Total International Flights Passengers	-0.759***
	(0.237)
Constant	-0.487
	(0.646)
Observations	237
R-squared	0.774

Short-Run ARDL Model Results	(10)
Variables	***Ln_Total International Flights Passengers***
L. Ln_Total International Flights Passengers	0.848***
	(0.0640)
L2. Ln_Total International Flights Passengers	-0.0213
	(0.0840)
L3. Ln_Total International Flights Passengers	0.149**
	(0.0637)
Ln_Total Registered Businesses	0.0617***
	(0.0172)
L1. Ln_Total Registered Businesses	-0.00433
	(0.0186)
L2. Ln_Total Registered Businesses	0.00165
	(0.0186)
L3. Ln_Total Registered Businesses	-0.0518***
	(0.0156)
Constant	0.253
	(0.174)
Observations	237
R-squared	0.954

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. *Source:* Authors' calculations.

The short-run estimates of ARDL show that *Total International Flights Passengers* in DIA affects *Total Registered Businesses in Denver MSA* in the short-run (in log form) – in 1-st lag. According to the test results, one percent point change in the 1-st lag of *Total Registered Businesses* is associated with about 0.75 percent point decrease in *Total Registered Businesses in Denver MSA*, on average, ceteris paribus.

The short-run estimates of ARDL show that *Total Registered Businesses in Denver MSA* affects *Total International Flights Passengers* in DIA in the short-run (in log form) – in 3-rd lag. According to the test results, one percent point change in the 3-rd lag of *Total Registered Businesses* is associated with about 0.05 percent point decrease in *Total International Flights Passengers* in DIA, on average, ceteris paribus.

In general, analyzing the pair of *Total Registered Businesses in Denver MSA* and *Total International Flights Passengers* demonstrated the following result in Table-2.11: both variables demonstrate bi-directional short-run causality, with higher magnitude effect from *Total International Flights Passengers* in DIA variable on *Total Registered Businesses* variable in Denver MSA.

Table-2.11. Pair-3 Results

Denver MSA Region		
Dependent variable	**Independent variable**	**Causality**
Total Registered Businesses in Denver MSA	*Total International Flights Passengers*	Short-run (lag-1), negative, significant
Total International Flights Passengers	*Total Registered Businesses in Denver MSA*	Short-run (lag-3), negative, significant

Source: Authors' elaborations.

Testing Pair: ***Total Registered Businesses in Denver MSA vs. Total Flights (DIA)***

The analysis of this pair of variables revealed no significant short-run/long-run relationship between both variables. Results of analysis are not included in the paper but are available upon request.

Testing Pair: ***Total Registered Businesses in Denver MSA vs. Total Cargo (DIA)***

The analysis of these pair of variables revealed: no significant short-run/long-run relationship between *Total Registered Businesses in Denver MSA* and *Total Cargo (DIA)*; inconclusive test results for reverse case of: *Total Cargo (DIA)* and *Total Registered Businesses in Denver MSA*. Results of analysis are not included in the paper but are available upon request.

2.4.3. VAR and VECM Model Results

The pair of variables, which demonstrated the same order of integration in ADF stationarity test are considered for further analysis of Granger causality through VAR or VECM modeling. In this case, the series might be cointegrated which, if not addressed, may result in spurious estimates.

The Johansen test is demonstrated with two similar test statistics: trace and maximum eigenvalue statistics. The null hypothesis in the trace statistics is existence k cointegrating equations, the alternative hypothesis is $k-1$ cointegrating relations. The maximum eigenvalue statistics tests for the null hypothesis of r cointegrating relations against the alternative of $r+1$. Both methods in general show similar decisions on the number of cointegration relations.

The analysis of the following variables revealed *zero cointegration equation* test results (no long-run relationship), and thus demonstrated no further Granger causality significant results after performing VAR modeling:

1) *Total Non-Farm Employment in Denver MSA* and *Total International Flights Passengers (DIA):* no Granger causality relationships between both variables, no autocorrelation detected.
2) *Total Non-Farm Employment in Colorado* and *Total International Flights Passengers (DIA):* no Granger causality relationships between both variables, autocorrelation detected.
3) *Total Non-Farm Employment in Denver MSA* and *Total Flights (DIA):* no Granger causality relationships between both variables, no autocorrelation detected.
4) *Total Non-Farm Employment in Colorado* and *Total Flights (DIA):* no Granger causality relationships between both variables, autocorrelation detected.

Results of analysis are not included in the paper but are available upon request.

Then, the next pair was analyzed and results are presented in Tables-2.12-2.14.

Testing Pair-4: *Total Non-Farm Employment in Denver MSA* and *Total Cargo (DIA)*

Table-2.12. Johansen Cointegration Test Results

Variables:	***Total Non-Farm Employment in Denver MSA and Total Cargo (DIA) Obs=236, Lags=4***				
Maximum rank	Parms	LL	Eigenvalue	Trace statistic	5% critical value
0	14	1531.19	.	11.288*	15.41
1	17	1536.56	0.044	0.5538	3.76
2	18	1536.84	0.002		

**Rejection of hypothesis at 5% significance level.*

Source: Authors' calculations.

Test statistics indicate that there exists a zero cointegration relationship between *Total Non-Farm Employment in Denver MSA* and *Total Cargo (DIA)*. That means that there exists a short-run relationship between these two variables, so we need to perform VAR modeling and the Granger causality approach to reveal short-run relationship between variables (results to be discussed in next section).

Testing Pair-5: *Total Non-Farm Employment in Colorado* and *Total Cargo (DIA)*

Table-2.13. Johansen Cointegration Test Results

Variables:	***Total Non-Farm Employment in Colorado and Total Cargo (DIA) Obs=236, Lags=4***				
Maximum rank	Parms	LL	Eigenvalue	Trace statistic	5% critical value
0	14	1555.15	.	10.2374*	15.41
1	17	1559.96	0.039	0.6148	3.76
2	18	1560.27	0.002		

**Rejection of hypothesis at 5% significance level.*

Source: Authors' calculations.

Test statistics indicate that there exists zero cointegration relationship between *Total Non-Farm Employment in Colorado* and *Total Cargo (DIA)*. That means that there exists a short-run relationship between these two variables, so we need to perform VAR modeling and the Granger causality approach to reveal short-run relationship between variables (results to be discussed in next section).

Testing Pair-6: *Total Registered Businesses in Denver MSA* and *Total Domestic Flights Passengers*

Table-2.14. Johansen Cointegration Test Results

Variables:	***Total Registered Businesses in Denver MSA and Total Domestic Flights Passengers Obs=236, Lags=4***				
Maximum rank	Parms	LL	Eigenvalue	Trace statistic	5% critical value
0	14	362.39	.	38.222	15.41
1	17	381.50	0.149	0.0050*	3.76
2	18	381.51	0.00002		

**Rejection of hypothesis at 5% significance level.*

Source: Authors' calculations.

Test statistics indicate that there exists at least one cointegration relationship between *Total Registered Businesses in Denver MSA* and *Total Domestic Flights Passengers (DIA)*. That means that there exists a long-run relationship between these two variables, so we need to perform the Vector Error Correction Model (VECM) approach to reveal long-run relationship between variables (results to be discussed in next section).

Testing Pair-7: *Total Registered Businesses in Denver MSA* and *Total Net Cargo (DIA)*

Table-2.15. Johansen Cointegration Test Results

Variables:	***Total Registered Businesses in Denver MSA and Total Net Cargo (DIA) Obs=236, Lags=4***				
Maximum rank	Parms	LL	Eigenvalue	Trace statistic	5% critical value
0	14	-28.32	.	16.5991	15.41
1	17	-21.12	0.059	2.2020*	3.76
2	18	-20.02	0.009		

**Rejection of hypothesis at 5% significance level.*

Source: Authors' calculations.

Test statistics indicate that there exists at least one cointegration relationship between *Total Registered Businesses in Denver MSA* and *Total Net Cargo (DIA)*. That means that there exists a long-run relationship between these two variables, so we need to perform the Vector Error Correction Model (VECM) approach to reveal long-run relationship between variables (results to be discussed in next section). This pair demonstrated no Granger causality between variables. The results are not included in the paper, but they are available from the authors.

Testing Pairs-4,5: *Total Non-Farm Employment (in Denver MSA and Colorado region)* and *Total Cargo (DIA)*

The results of the cointegration test suggest that there exists short-run relationship between *Total Non-Farm Employment* (both in Denver MSA and Colorado region) and *Total Cargo (DIA)* variables, but it does not indicate the direction of the causal relationship. Since zero cointegration is detected, the Vector Autoregressive Model (VAR) is used (VECM is used in the presence of cointegration), which tests for short-run causality relationship between the series. The results of VAR analysis are reported in Table-2.16.

Table-2.16. VAR Model Results

Denver MSA Region

Short-run VAR Model Results	(11)	(12)
Variables	***Ln_ Total Non-Farm Employment in Denver MSA_d1***	***Ln_Total Cargo in DIA_d1***
L. Ln_ Total Non-Farm Employment in Denver MSA_d1	1.186***	0.414
	(0.0626)	(1.754)
L2. Ln_ Total Non-Farm Employment in Denver MSA_d1	0.0811	-2.357
	(0.100)	(2.809)
L3. Ln_ Total Non-Farm Employment in Denver MSA_d1	0.0535	2.054

	(0.0999)	(2.801)
L4. Ln_ Total Non-Farm Employment in Denver MSA_d1	-0.321***	-0.0609
	(0.0619)	(1.737)
L. Ln_Total Cargo in DIA_d1	0.00787***	0.361***
	(0.00231)	(0.0648)
L2. Ln_Total Cargo in DIA_d1	-0.00201	0.429***
	(0.00221)	(0.0619)
L3. Ln_Total Cargo in DIA_d1	0.00158	0.439***
	(0.00224)	(0.0627)
L4. Ln_Total Cargo in DIA_d1	-0.00843***	-0.280***
	(0.00223)	(0.0626)
Constant	0.0203	0.535
	(0.0180)	(0.504)
Observations	236	236

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Colorado Region

Short-run VAR Model Results	(13)	(14)
Variables	***Ln_ Total Non-Farm Employment in Colorado_d1***	***Ln_Total Cargo in DIA_d1***
L. Ln_ Total Non-Farm Employment in Colorado_d1	0.328***	0.365
	(0.0603)	(1.897)
L2. Ln_ Total Non-Farm Employment in Colorado_d1	0.429***	0.604
	(0.0588)	(1.848)
L. Ln_Total Cargo in DIA_d1	0.00508**	-0.700***
	(0.00201)	(0.0631)
L2. Ln_Total Cargo in DIA_d1	-0.00105	-0.361***
	(0.00199)	(0.0626)
Constant	0.000242*	-0.00469
	(0.000124)	(0.00391)
Observations	237	237

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Source: Authors' calculations.

The estimates of VAR model show that *Total Cargo (DIA)* affects *Total Non-Farm Employment* both in Denver MSA and Colorado region in the short-run (in log form). According to the test results for Denver MSA: In Lag-1: one percent point change in the *Total Cargo (DIA)* is associated with about 0.007 percent point *increase* in *Total Non-Farm Employment in Denver MSA*, on average, ceteris paribus. In Lag-4: one percent point change in the *Total Cargo (DIA)* is associated with about 0.008 percent point *decrease* in *Total Non-Farm Employment in Denver MSA*, on average, ceteris paribus.

According to the test results for the Colorado region: one percent point change in the *Total Cargo (DIA)* is associated with about 0.005 percent point *increase* in *Total Non-Farm Employment in Colorado*, on average, ceteris paribus. On the other hand, in the reverse case, where *Total Cargo (DIA)* is defined as dependent variable, the estimates of VAR show that *Total Non-Farm Employment* both in Denver MSA and Colorado region does not affect *Total Cargo (DIA)* in the short-run (in log form).

After obtaining the VAR model results the Granger causality test was performed to determine the direction of causality. Then Lagrange Multiplier test for checking the autocorrelation between variables is performed. Results of Granger Causality and autocorrelation test are presented in Table-2.17.

Table-2.17. Granger causality Wald tests Results

Denver MSA Region

Equation	**Excluded**	**chi2**	**df**	**Prob > chi2**
Ln_ Total Non-Farm Employment in Denver MSA	*Ln_Total Cargo in DIA*	24.849	4	0.000
Ln_ Total Non-Farm Employment in Denver MSA	*ALL*	24.849	4	0.000
Ln_Total Cargo in DIA	*Ln_Total Non-Farm Employment in Denver MSA*	2.5542	4	0.635
Ln_Total Cargo in DIA	*ALL*	2.5542	4	0.635

Colorado Region

Equation	**Excluded**	**chi2**	**df**	**Prob > chi2**
Ln_Total Non-Farm Employment in Colorado	*Ln_Total Cargo in DIA*	11.495	2	0.003
Ln_Total Non-Farm Employment in Colorado	*ALL*	11.495	2	0.003

Ln_ Total Cargo in DIA	*Ln_Total Non-Farm Employment in Colorado*	0.358	2	0.836
Ln_ Total Cargo in DIA	*ALL*	0.358	2	0.836

Lagrange-Multiplier Test

Denver MSA Region				Colorado Region			
Lag	**chi2**	**df**	**Prob > chi2**	**Lag**	**chi2**	**df**	**Prob > chi2**
1	6.1892	4	0.185	1	39.309	4	0.00000
2	7.9656	4	0.092	2	31.511	4	0.00000

H_0: no autocorrelation at lag order. Source: Authors' calculations.

The results of Granger causality test reveal the direction: there is a significant short-run effect of *Total Cargo* in DIA on *Total Non-Farm Employment* both in Denver MSA and Colorado region, which proved the above VAR modeling results. The results of the LM test reveal that model is specified well: there is no evidence of autocorrelation between variables.

In general, analyzing the pair of *Total Cargo* in DIA and *Total Non-Farm Employment* both in Denver MSA and Colorado demonstrated the following result in Table-2.18:

Table-2.18. Pairs-4,5 Results

Denver MSA Region		
Dependent variable	**Independent variable**	**Causality**
Total Non-Farm Employment in Denver MSA	*Total Cargo in DIA*	Short-run (lags-1, 4), positive/negative, significant
Colorado Region		
Dependent variable	**Independent variable**	**Causality**
Total Non-Farm Employment in Colorado	*Total Cargo in DIA*	Short-run, positive, significant

Source: Authors' elaborations.

In reverse case, where we analyzed the impact of *Total Non-Farm Employment* both in Denver MSA and Colorado region on *Total Cargo in DIA,* revealed no Granger causality. The results are not included in the paper, but they are available from the authors.

Testing Pair-6: *Total Registered Businesses in Denver MSA* and *Total Domestic Flights Passengers*

The results of the cointegration test suggest that there exists a long-run relationship between *Total Registered Businesses in Denver MSA* and *Total Domestic Flights Passengers* variables, but it does not indicate the direction of the causal relationship. Since cointegration is detected, the Vector Error Correction Model (VECM) is used (VAR is unspecified in the presence of cointegration), which tests for both short-run and long-run causality relationship between the series. The results of VECM analysis are reported in Table-2.19.

Table-2.19. VECM Model Results

Long-run VECM Model Results	(15)
Variables	***Ln_ Total Registered Businesses in Denver MSA***
L._ce1	-0.417***
	(0.0893)
LD. Ln_Total Registered Businesses	-0.277***
	(0.0870)
L2D. Ln_Total Registered Businesses	-0.0528
	(0.0883)
L3D. Ln_Total Registered Businesses	-0.0243
	(0.0933)
L4D. Ln_Total Registered Businesses	-0.0573
	(0.0904)
L5D. Ln_Total Registered Businesses	0.0555
	(0.0824)
L6D. Ln_Total Registered Businesses	-0.142**
	(0.0602)
LD. Ln_Total Domestic Flights Passengers	-1.209**
	(0.506)
L2D. Ln_Total Domestic Flights Passengers	-1.172**
	(0.516)

Long-run VECM Model Results	(16)
Variables	***Ln_Total Domestic Flights Passengers in DIA***
L._ce1	-0.120***
	(0.0388)
LD. Ln_Total Domestic Flights Passengers	-0.338***
	(0.0718)
L2D. Ln_Total Domestic Flights Passengers	-0.186**
	(0.0733)
L3D. Ln_Total Domestic Flights Passengers	-0.126*
	(0.0738)
L4D.Ln_Total Domestic Flights Passengers	-0.104
	(0.0733)
L5D.Ln_Total Domestic Flights Passengers	-0.123*
	(0.0711)
L6D. Ln_Total Domestic Flights Passengers	-0.0422
	(0.0656)
LD. Ln_Total Registered Businesses	-0.0701***
	(0.0124)
L2D. Ln_Total Registered Businesses	-0.0621***

L3D. Ln_Total Domestic Flights Passengers	-0.623
	(0.520)
L4D. Ln_Total Domestic Flights Passengers	-1.077**
	(0.516)
L5D. Ln_Total Domestic Flights Passengers	-0.291
	(0.501)
L6D. Ln_Total Domestic Flights Passengers	1.978***
	(0.462)
Constant	0.000972
	(0.0197)
Observations	233

Johansen normalization restriction imposed

_ce1 Ln_Total Domestic Flights Passengers	-3.0597***
	(0.2088)

	(0.0125)
L3D. Ln_Total Registered Businesses	-0.0450***
	(0.0132)
L4D. Ln_Total Registered Businesses	-0.0458***
	(0.0128)
L5D. Ln_Total Registered Businesses	-0.0556***
	(0.0117)
L6D.Ln_Total Registered Businesses	-0.0156*
	(0.00855)
Constant	0.0103***
	(0.00280)
Observations	233

Johansen normalization restriction imposed

_ce1 Ln_ Total Registered Businesses	- 0.3268 ***
	(0.0214)

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Source: Authors' calculations.

The estimates of VECM show that *Total Domestic Flights Passengers* affects *Total Registered Businesses in Denver MSA* in the long run (in log form). According to the test results, a one percent point change in the *Total Domestic Flights Passengers* is associated with about 3.0 percent point increase in *Total Registered Businesses* in Denver MSA, on average, ceteris paribus. But in the short-run (lags) *Total Domestic Flights Passengers* demonstrated the negative effect on *Total Registered Businesses*. The adjustment term shows that errors in the previous period will be corrected in the current period, the coefficient is negative (-0.417) and statistically significant at 5% level.

On the other hand, in the reverse case, the estimates of VECM show that *Total Registered Businesses in Denver MSA* affects *Total Domestic Flights Passengers* in DIA in the long run (in log form). According to the test results, one percent point change in the *Total Registered Businesses* is associated with about 0.32 percent point increase in *Total Domestic Flights*

Passengers in DIA, on average, ceteris paribus. In the short run (lags) *Total Registered Businesses* also demonstrated the negative effect on *Total Domestic Flights Passengers*. The adjustment term shows that errors in the previous period will be corrected in the current period, the coefficient is negative (-0.120) and statistically significant at the 5% level.

To estimate the validity of the estimated models, diagnostic tests were performed and results are presented in *Appendix-B*.

In general, analyzing the pair of *Total Registered Businesses in Denver MSA* and *Total Domestic Flights Passengers* demonstrated the following results in Table-2.20: both variables demonstrate bi-directional long-run causality, with higher magnitude effect from *Total Domestic Flights Passengers* at DIA variable on *Total Registered Businesses in Denver MSA*.

Table-2.20. Pair-6 Results

Denver MSA Region		
Dependent variable	**Independent variable**	**Causality**
Total Registered Businesses in Denver MSA	*Total Domestic Flights Passengers in DIA*	Long-run, positive, significant, higher magnitude effect. Short-run, negative, significant
Total Domestic Flights Passengers in DIA	*Total Registered Businesses in Denver MSA*	Long-run, positive, significant, lower magnitude effect. Short-run, negative, significant

Source: Authors' elaborations.

Another pair that revealed long-term relationship in cointegration test above between *Total Registered Businesses in Denver MSA* and *Total Net Cargo (DIA)* demonstrated no Granger causality between variables. The results are not included in the paper, but they are available upon request.

2.4.4. Impulse Responses Functions Results of VAR and VECM Model

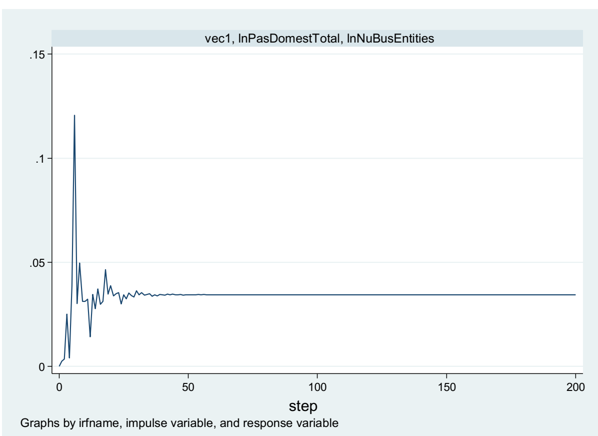
Impulse Responses Functions (IRFs) are the most commonly applied tools for describing dynamic reactions in vector autoregressive (VAR) and vector error-correction models (VECM) analysis[4]. The impulse response function (IRF) is used to reflect the response of an endogenous variable on the impact of innovation (shock) in an exogenous viable. Specifically, if a standard deviation impact is given to the random error term, it will affect the current and future values of the endogenous variables.

Whereas IRFs from a stationary VAR die out over time, IRFs from a cointegrating VECM do not always die out. Because each variable in a stationary VAR has a time-invariant mean and finite, time-invariant variance, the effect of a shock on any one of these variables must die out so that the variable can revert to its mean. In contrast, the I (1) variables modeled in a cointegrating VECM are not mean reverting, and the unit moduli in the companion matrix implies that the effects of some shocks will not die out over time. These two possibilities gave rise to new terms. When the effect of a shock dies out over time, the shock is said to be *transitory*. When the effect of a shock does not die out over time, the shock is said to be *permanent*.

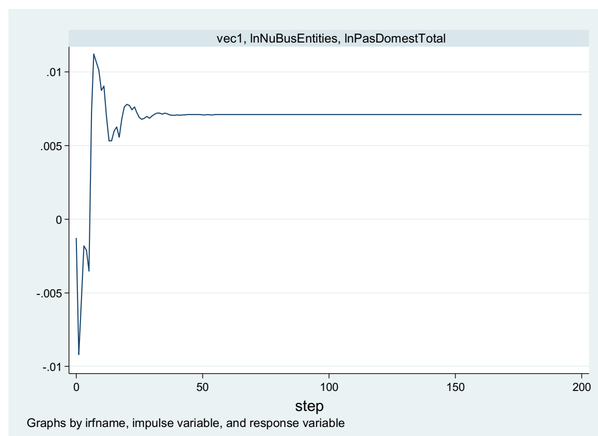
Figure-2.2 shows the impulse response from VECM model: responses of *Total Domestic Flights Passengers in DIA* and *Total Registered Businesses in Denver MSA* to an orthogonalized one S.D. innovation. The blue line is the impulse response estimate for the period, and the border of highlighted lines are the one-standard error confidence bands.

[4] To address some of the limitations of traditional VAR and VECM approaches—such as sensitivity to model specification and the requirement of system-wide estimation—in the further research of this study it is recommended to employ the *Local Projections* method developed by Jordà (2005). LPs offer a more robust and flexible framework for estimating impulse responses, as they allow for direct estimation at each forecast horizon and are less prone to propagation bias and misspecification.

The graphs indicate that an orthogonalized shock to the *Total Domestic Flights Passengers in DIA* has a *permanent effect* on the *Total Registered Businesses in Denver MSA* with a higher magnitude (Fig-2.2a). The effect fluctuated in the first 40 periods. An orthogonalized shock to the average of *Total Registered Businesses in Denver MSA* has a *permanent effect* on the average of *Total Domestic Flights Passengers in DIA* with a lower magnitude (Fig-2.2b). The effect fluctuated in the first 50 periods. According to this model, unexpected shocks that are local to *Total Domestic Flights Passengers in DIA* will have a permanent higher magnitude effect on the *Total Registered Businesses in Denver MSA*, but unexpected shocks that are local to the *Total Registered Businesses in Denver MSA* will have a permanent lower magnitude effect on the *Total Domestic Flights Passengers in DIA*.



a) Impulse Response Function of *Total Domestic Flights Passengers in DIA* on impact from *Total Registered Businesses in Denver MSA.*



b) Impulse Response Function of *Total Registered Businesses in Denver MSA* on impact from *Total Domestic Flights Passengers in DIA.*

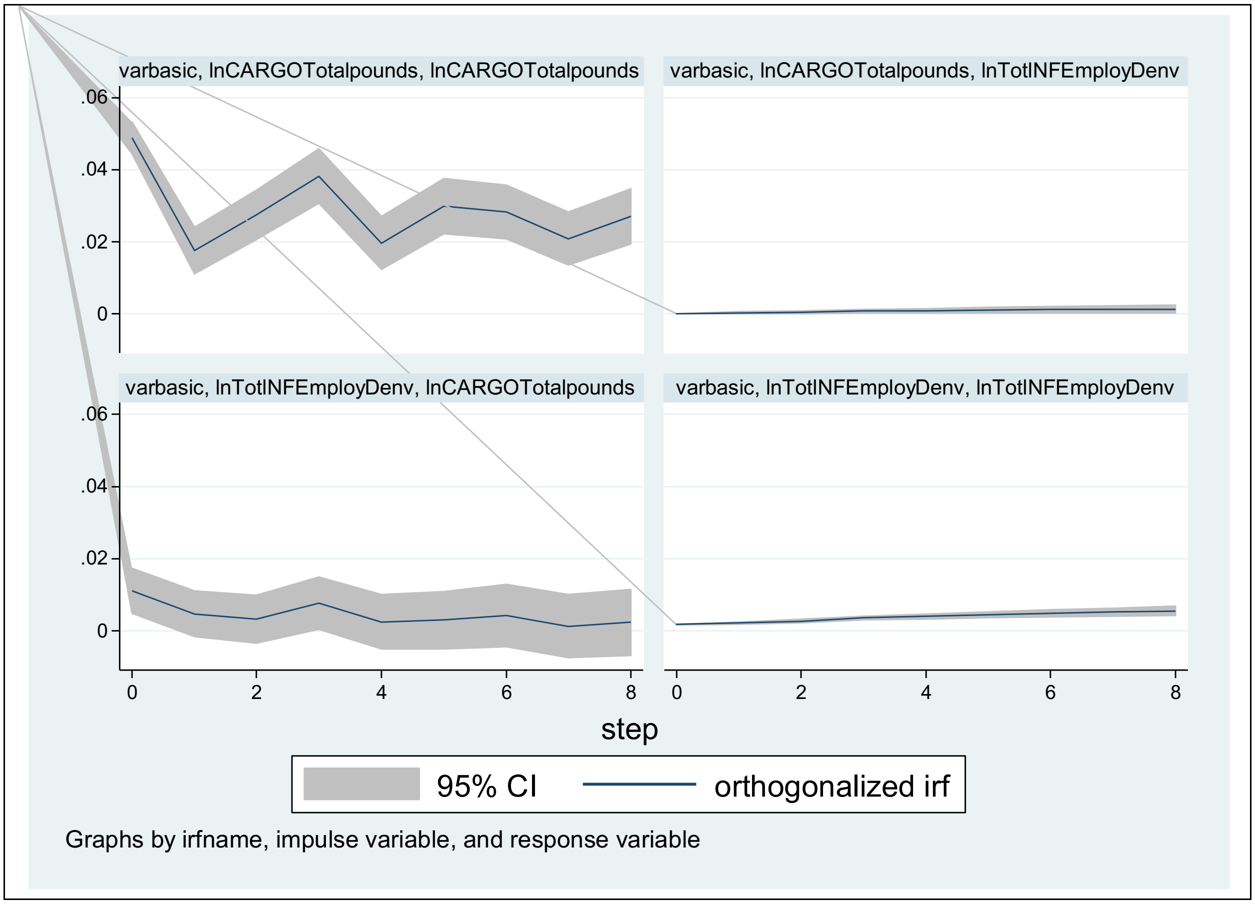
Figure-2.2. Impulse Response Functions of *Total Domestic Flights Passengers (DIA) and Total Registered Businesses (Denver MSA).*

Source: Authors' calculations.

From Figure-2.3a, when there is an innovation on *Total Non-Farm Employment (Denver MSA)* for horizon periods H = 8, the response of *Total Cargo (DIA)* shows persistent and minimal positive increase in short-term development. From Figure-2.3b, when there is an innovation on *Total Cargo (DIA),* the response of *Total Non-Farm Employment (Denver MSA)* fluctuated a little in the beginning 8 periods and then stabilized with a minimal positive effect over short-term development (but with a higher magnitude).

These results indicated that the mutual positive impact of *Total Cargo (DIA)* and *Total Non-Farm Employment (Denver MSA)* existed for a short-period, and effect of *Total Cargo* was more sensitive on *Total Non-Farm Employment* (more fluctuations), which proved above results from VAR model.

Figure-2.3. Impulse Response Functions of *Total Cargo (DIA)* and *Total Non-Farm Employment (Denver MSA)*



Source: Authors' calculations.

2.5. Discussion

The discussion is divided into sections based on the regions, with findings for major variables and their comparing.

2.5.1. Results in Denver MSA Region

The summary of all tested casual relationships in Denver MSA region is presented in Table-2.21.

Table-2.21. Causality Results for Denver MSA Region

ARDL Model result		
Dependent variable	**Independent variable**	**Causality**
Total Non-Farm Employment	*Total Domestic Flights Passengers*	Short-run, positive, significant, lower magnitude effect
Total Domestic Flights Passengers	*Total Non-Farm Employment*	Short-run (lags-3,4), positive/negative, significant, higher magnitude effect
Dependent variable	**Independent variable**	**Causality**
Total Non-Farm Employment	*Difference of Inbound-Outbound Cargo (DIA)*	Short-run (lag-2), negative, significant, lower magnitude effect
Difference of Inbound-Outbound Cargo (DIA)	*Total Non-Farm Employment*	Long-run, positive, significant Short-run (lag-2), negative, significant, higher magnitude effect
Dependent variable	**Independent variable**	**Causality**
Total Registered Businesses in Denver MSA	*Total International Flights Passengers*	Short-run (lag-1), negative, significant, higher magnitude effect
Total International Flights Passengers	*Total Registered Businesses in Denver MSA*	Short-run (lag-3), negative, significant, lower magnitude effect
VAR Model result		
Dependent variable	**Independent variable**	**Causality**
Total Non-Farm Employment in Denver MSA	*Total Cargo in DIA*	Short-run (lags-1, 4), positive/negative, significant
VECM Model result		
Dependent variable	**Independent variable**	**Causality**
Total Registered Businesses in Denver MSA	*Total Domestic Flights Passengers in DIA*	Long-run, positive, significant, higher magnitude effect. Short-run, negative, significant
Total Domestic Flights Passengers in DIA	*Total Registered Businesses in Denver MSA*	Long-run, positive, significant, lower magnitude effect. Short-run, negative, significant

Source: Authors' elaborations

- **Effect of Air Transportation on Regional Development in Denver MSA**

The results demonstrated that there is a clear bi-directional long-run Granger causality between *Total Registered Businesses in Denver MSA* and *Total Domestic Flights Passengers in DIA,* with higher magnitude effect from *Total Domestic Flights Passengers*. This result showed that even though business entities and domestic flights are important to each other and for overall regional development, airport transportation brings higher effect to the economic development of the regions.

Also, results show that *Total Cargo* operated in DIA has short-run alternate effects on *Total Non-Farm Employment* in Denver MSA. This result reveals a normal business cycle activity trend in the region and confirms that local employment is hugely affected by cargo transportation to the region due to possible intense business relations.

Lastly, results show that *Registered Businesses* in Denver MSA region and *International Flights* operated in DIA have short-run bi-directional negative effects on each other, with higher magnitude effect from *Total International Flights Passengers*. This result reveals that international flights affect businesses stronger negatively in the short-run plausibly due to a normal business cycle.

- **Effect of Regional Development in Denver MSA on Air Transportation**

On the other hand, *Total Non-Farm Employment* in Denver MSA demonstrated short-term bi-direction positive/negative effect on *Total Domestic Flights Passengers* in Denver MSA, with higher magnitude effect from *Total Non-Farm Employment,* which shows that local economic development is sometimes a stronger driver for local air transportation system than vice-versa.

Another result shows that *Total Non-Farm Employment* is important for *Net Cargo in DIA* in the long run. This result confirms the hypothesis that the higher the employment in area, the higher the economic activity in the region will be and the more cargo will be brought to the region through different business channels.

2.5.2. Results in Colorado Region

The summary of all tested casual relationships in Colorado region is presented in Table-2.22.

Table-2.22. Causality Results for Colorado Region

ARDL Model result		
Dependent variable	**Independent variable**	**Causality**
Total Non-Farm Employment	*Total Domestic Flights Passengers*	Short-run, negative (Lag-3), significant, lower magnitude effect
Total Domestic Flights Passengers	*Total Non-Farm Employment*	Short-run, positive, significant, higher magnitude effect
Dependent variable	**Independent variable**	**Causality**
Total Non-Farm Employment	*Difference of Inbound-Outbound Cargo (DIA)*	Short-run, negative (Lag-2), significant, lower magnitude effect
Difference of Inbound-Outbound Cargo (DIA)	*Total Non-Farm Employment in Colorado*	Long-run, positive, significant Short-run, positive, significant, higher magnitude effect
VAR Model result		
Dependent variable	**Independent variable**	**Causality**
Total Non-Farm Employment in Colorado	*Total Crago in DIA*	Short-run, positive, significant

Source: Authors' elaborations

- **Effect of Regional Development in Colorado region on Air Transportation**

There is a bi-directional short-term impact between *Total Non-Farm Employment* in Colorado *and Total Domestic Flights Passengers (DIA)*: positive effect between employment on domestic passengers, and negative between domestic passengers on employment (Lag-3), with higher magnitude effect from *Total Non-Farm Employment*. This demonstrates how local business employment outlook increases the domestic flights passengers' traffic in whole Colorado.

Another result shows that *Total Non-Farm Employment* in Colorado is important for *Net Cargo in DIA* in the long run, with higher magnitude effect from *Total Non-Farm Employment*. This result confirms the hypothesis that the higher the employment in area, the higher the economic activity in the region will be and the more cargo will be brought to the region through different business channels.

- **Effect of Air Transportation on Regional Development in Colorado region**

Additionally, the results show that *Total Cargo* transported in DIA has a short-run positive effect on *Total Non-Farm Employment* in Colorado. This result confirms that local employment in the whole of Colorado is affected by local business activity. In sum, higher business activity is associated with more cargo and then with higher employment in turn. Also, the results demonstrate the short-run negative effect of *Net Cargo in DIA* on *Total Non-Farm Employment* in Colorado (Lag-2), which might be the result of business cycles.

2.5.3. Comparing Denver MSA and Colorado Regions Results

When we compare Colorado results to Denver MSA results, we can draw several conclusions. The effect of Domestic Flight Passengers on the whole of Colorado employment is negative in the short-run (lag-3), while it had positive short-run effect in Denver MSA. This might be because domestic passengers have more business connections in Denver MSA region, rather than in more remote cities in the whole of Colorado. More far areas from Denver – business activity is more fading out.

There is similar short-run negative effect of *Net Cargo* on total Colorado employment, as it was in Denver (lag-2) – possibly because of labor productivity: higher is difference between inputs/outputs (in absolute terms) – the less employees are needed for production process (when in short-run capital intensity is fixed).

In the reverse case, there is a similar result for Colorado as it was in Denver: a strong positive effect of *Total Employment* on *Total Net Cargo* both in the long-run and in the short-run. The more employment is in the area, the higher is *Net Cargo*.

When we focus on cargo, we have similar short-run positive effects of *Total Cargo* on *Total Employment* in Colorado, as it is in Denver: this confirms that overall, more cargo, possibly because of higher business activity, brings more employment to areas.

Overall, we can say that Colorado results are similar to the results from Denver MSA, except for domestic passengers effect, and that air transportation and regional development variables demonstrate mostly reverse causality, as predicted by the literature.

2.6. Conclusions

The paper examines the dynamic causal relationship among the series of air transportation and regional economic development variables for the Denver MSA in comparison to Colorado overall for the period 2000-2019. The research objective is to estimate if there is a Granger causality between air transportation and regional development (and vice-versa). The research question is: what is the effect of air transportation in Denver International Airport (DIA) on the emergence and growth of businesses in the Denver MSA and on the change of employment in both Denver and the state of Colorado. We focus our analysis on the local area of Denver versus whole Colorado region – to see the difference of effect in a business hub like Denver and more remote areas included into overall Colorado region.

Time series variables first are checked for stationarity using Augmented Dickey-Fuller Test for Unit Root. Since several of the variables in the dataset are of mixed integration order (i.e., stationary and non-stationary variables), the standard Johansen test and VAR/VECM modelling procedure are not appropriate for these cases.

Instead, the ARDL modeling approach is used to distinguish between short-run and long-run effects between tested variables. It was revealed that *Total Non-Farm Employment*, both in Denver and statewide in Colorado, has a uni-directional, positive long-run effect on *Net Cargo* processed in DIA, which confirms the hypothesis that higher is economic activity in the region (proxied by employment) then more cargo will be brought to the region. ARDL modelling also demonstrated short-run, negative sign bi-direction casualties from *Total Business Entities in Denver* to *Total International Flights in DIA*. This reflects the usual business activity trend along with the fact that Denver is a regional hub where many companies engage in international travel.

The last result from ARDL modeling revealed the short-run, positive sign causality from *Domestic Flight Passengers* to *Total Employment* in Denver, but negative short-run effect on *Total Employment* in Colorado. This result also aligns with the hypothesis that domestic air transportation is important for local employment in the region, especially in a big hub city like Denver, and the effect fades out while spreading in more remote regions. *Total Cargo* in DIA affected Total Employment in Denver with alternate sign in the short-run and has a short-run positive effect of *Total Cargo* on *Total Employment* in whole Colorado, which suggests that more cargo, possibly because of higher business activity, brings more employment to regional areas.

The two remaining pairs of stationary variables (*Total Domestic Flight Passengers* and *Total Business Entities in Denver MSA*) are chosen for further analysis in a traditional Granger causality framework. The Johansen cointegration test was performed and at least one cointegrating equation has been detected between these variables in the long run. The result of VECM model estimation revealed bi-directional, positive, significant causality between *Total Domestic Flight*

Passengers and *Total Business Entities* in Denver MSA. The magnitude of effect is higher when *Total Domestic Passengers* is defined as an independent variable. So, air transportation has a larger effect on regional development.

Impulse Responses Functions (IRFs) are used to describe dynamic reactions in vector autoregressive (VAR) and vector error-correction models (VECM). According to the results, unexpected shocks that are local to *Total Domestic Flights Passengers in DIA* will have a permanent higher magnitude effect on the *Total Registered Businesses in Denver MSA*, but unexpected shocks that are local to the *Total Registered Businesses in Denver MSA* will have a permanent lower magnitude effect on the *Total Domestic Flights Passengers in DIA,* which proves the above results from VECM model.

The policy implication of these results is that air transportation plays an important role in both local and regional development, so to develop remote regions, government needs to promote the creation/operation of airports in the regions, including more routes, because it will lead to increase of economic activity in the regions. According to the results, a one percent point change in the *Total Domestic Flights Passengers* is associated with about 3.0 percent point increase in *Total Registered Businesses* in Denver MSA, on average, ceteris paribus. This confirms that domestic routes are important factors in the emergence of local businesses in the Denver MSA area. The effect of *Domestic Flight Passengers* on Denver MSA employment is also positive, while it is negative for whole Colorado region in the short-run: one percent point change in the *Total Domestic Flights Passengers* brings a change of about 0.006 percent point increase in *Total Non-Farm Employment* in Denver MSA and about 0.004 percent point decrease in *Total Non-Farm Employment* in Colorado region, on average, ceteris paribus. This result also shows how important domestic flights for employment are in local areas. More domestic passengers bring higher employment in the hub area, and this effect fades out in more remote areas.

In addition, *Total Cargo* positively affects employment both in the Denver MSA and the overall Colorado region. According to the results, a one percent point change in the *Total Cargo (DIA)* is associated with about 0.007 percent point increase in *Total Non-Farm Employment in Denver MSA* and with about 0.005 percent point increase in *Total Non-Farm Employment in Colorado*, on average, ceteris paribus. These results are proved by using Impulse Responses Functions for above VAR model, which indicate that the mutual positive impact of *Total Cargo (DIA)* and *Total Non-Farm Employment (Denver MSA)* existed for a short-period, and the effect of *Total Cargo* is more sensitive on *Total Non-Farm Employment* (more fluctuations). Overall, we see a positive effect of air transportation on local economic development in the case of Denver MSA and Colorado regions. The positive employment effect in Denver and Colorado is likely driven by the wide economic reach of Denver International Airport. As a major hub, DIA creates jobs directly through airport operations and airline activity, but it also supports jobs indirectly in areas like logistics, hospitality, and trade. Its central location helps attract businesses that rely on fast connections, including e-commerce and professional services. The airport has also encouraged development in nearby areas, boosting construction and service jobs. Overall, DIA strengthens Colorado's link to national and global markets, which helps drive job growth across multiple sectors.

There are some caveats to be noted. Denver International Airport is a transportation hub that connects different cities. People might travel to/from Denver not only for visiting Colorado for business, but also for transferring purposes to go to other places beside Colorado. As a result, transferring passengers might play a considerable but different role in local economic development. Data limitations prevented detailed study of these nuances. For example, authors attempted to find the data on transfer passengers – to be able to include it as a control to our models of emerging businesses or employment in Denver MSA or Colorado region. Unfortunately, the full data on transfer passengers is not available by Denver International Airport for all months of 2000-2019 period. Conceptually, this opens the possibility for further research direction upon the availability of necessary data in the future.

As another limitation, it should also be noted that the present Colorado-based case study does not explicitly account for potential spillover effects to neighboring states that also utilize Denver International Airport (DEN), which may lead to an underestimation of the broader regional impacts.

REFERENCE:

1. Appold, S. J. (2015). The impact of airports on US urban employment distribution. *Environment and Planning A, 47*(2), 408-426. https://doi.org/10.1068/a130114p
2. Appold, S. J., & Kasarda, J. D. (2012). The airport city phenomenon: Evidence from large US airports. *Urban Studies, 50*(6), 1237-1259. https://doi.org/10.1177/0042098012464401
3. Bai, Y., & Wu, C.-L. (2022). The causality analysis of airports and regional economy: Empirical evidence from Jiangsu Province in China. *Sustainability, 14*(7), 4295. https://doi.org/10.3390/su14074295
4. Bel, G., & Fageda, X. (2008). Getting there fast: Globalization, intercontinental flights, and the location of headquarters. *Journal of Economic Geography, 8*(4), 471–495. https://doi.org/10.1093/jeg/lbn017
5. Blonigen, B. A., & Cristea, A. D. (2015). Air service and urban growth: Evidence from a quasi-natural policy experiment. *Journal of Urban Economics, 86*(1), 140-155. https://doi.org/10.1016/j.jue.2015.02.001
6. Brueckner, J. K. (2002). Airport congestion when carriers have market power. *American Economic Review, 92*(5), 1357–1375. https://doi.org/10.1257/000282802762024548
7. Button, K. J., Doh, S., & Yuan, J. (2009). The role of small airports in economic development. In *Acanços do Transporte Aéreo Brasileiro* (pp. 643-657). EPUSP. https://www.researchgate.net/publication/251864448_The_role_of_small_airports_in_economic_development
8. Button, K., & Lall, S. (1999). The economics of being an airport hub city. *Research in Transportation Economics, 5*, 75-105. https://doi.org/10.1016/S0739-8859(99)80005-5
9. Button, K., & Taylor, S. (2000). International air transportation and economic development. Journal of Air Transport Management, 6(4), 209-222. https://doi.org/10.1016/S0969-6997(00)00015-6
10. Chen, X., Xuan, C., & Qiu, R. (2021). Understanding spatial spillover effects of airports on economic development: New evidence from China's hub airports. *Transportation Research Part A: Policy and Practice, 143*(C), 48-60. https://doi.org/10.1016/j.tra.2020.11.013
11. Cidell, J. (2006). Air transportation, airports, and the discourses and practices of globalization. *Urban Geography, 27*(7), 651–663. https://doi.org/10.2747/0272-3638.27.7.651
12. Colorado Information Marketplace. (2024, November 28). *Colorado business entities (corporations, LLCs, etc.) registered with the Colorado Department of State (CDOS) since 1864*. Colorado Information Marketplace. Retrieved November 25, 2024, from https://data.colorado.gov/stories/s/pbek-aaa3
13. Denver International Airport. (n.d.). *Governance, reports, and financials*. Denver International Airport. Retrieved November 25, 2024, from https://www.flydenver.com/about-den/governance/reports-and-financials/
14. Dickey, D. A., & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica, 49*(4), 1057-1072. https://doi.org/10.2307/1912517
15. Dickey, D. A., Jansen, D. W., & Thornton, D. L. (1991). A primer on cointegration with an application to money and income. *Review*, Federal Reserve Bank of St. Louis, *March*, 58-78. https://doi.org/10.20955/r.73.58-78
16. Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, *55*(2), 251-276. https://doi.org/10.2307/1913236

17. Florida, R. (2012, May 23). Airports and the wealth of cities. *The Atlantic*. Retrieved from [https://aviationacrossamerica.org/news/2012/05/29/airports-and-the-wealth-of-cities/]
18. Forsyth, P., Gillen, D., Muller, J., & Niemeier, H. M. (2016). *Airport competition: The European experience*. Routledge.
19. Graham, A. (2014). *Managing airports: An international perspective* (4th ed.). Routledge. https://doi.org/10.4324/9780080473284
20. Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica, 37*(3), 424-438. https://doi.org/10.2307/1912791
21. Granger, C. W. J., & Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics, 2*(2), 111-120. https://doi.org/10.1016/0304-4076(74)90034-7
22. Green, R. K. (2007). Airports and economic development. *Journal of Regional Science, 47*(1), 1-22. https://doi.org/10.1111/j.1540-6229.2007.00183.x
23. Goetz, A. R. (1992). Air passenger transportation and growth in the U.S. urban system, 1950–1987. *Urban Studies, 29*(2), 265-278. https://doi.org/10.1111/j.1468-2257.1992.tb00580.x
24. Jordà, Ò. (2005). *Estimation and Inference of Impulse Responses by Local Projections*. American Economic Review, 95(1), 161–182. https://doi.org/10.1257/0002828053828518
25. Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12(2–3), 231–254. https://doi.org/10.1016/0165-1889(88)90041-3
26. Kasarda, J. D. (2000). Logistics & the rise of aerotropolis. *Real Estate Issues, 25*(4), 43. Gale Academic OneFile. Retrieved from https://link.gale.com/apps/doc/A71837302/AONE?u=anon~660aef65&sid=googleScholar&xid=f35aec47
27. Ke, M., Baker, D., & Collis, C. (2023). Impact of airports on landside industrial development: A case study of Brisbane Airport. *Land, 12*(7), 1327. https://doi.org/10.3390/land12071327
28. Lenaerts, B., Allroggen, F., & Malina, R. (2021). The economic impact of aviation: A review on the role of market access. *Journal of Air Transport Management, 91*(5), 102000. https://doi.org/10.1016/j.jairtraman.2020.102000
29. Lütkepohl, H. (2010). Impulse response function. In: Durlauf, S.N., Blume, L.E. (eds) Macroeconometrics and Time Series Analysis. The New Palgrave Economics Collection. Palgrave Macmillan, London. https://doi.org/10.1057/9780230280830_16
30. Metro Denver EDC. (2023). *Exports*. Metro Denver Economic Development Corporation. Retrieved from https://www.metrodenver.org/do-business/international-trade/exports
31. Mukkala, K., & Tervo, H. (2013). Air transportation and regional growth: Which way does the causality run? Environment and Planning A, 45(6), 1508-1520. https://doi.org/10.1068/a45298
32. Percoco, M. (2010). Airport Activity and Local Development: Evidence from Italy. Urban Studies, 47, 2427-2443. https://doi.org/10.1177/0042098009357966
33. Pesaran, H., & Shin, Y. (1995). An autoregressive distributed lag modeling approach to co-integration analysis. *Cambridge University Press*. https://doi.org/10.1017/CCOL0521633230.011
34. Pesaran, M. H., Shin, Y., & Smith, R. J. (2001). Bounds testing approaches to the analysis of level relationships. *Journal of Applied Econometrics*, *16*(3), 289-326. https://doi.org/10.1002/jae.616
35. Rao, B. B. (2007). Estimating short and long-run relationships: A guide for the applied economist. *Applied Economics, 39*(13), 1613-1625. https://doi.org/10.1080/00036840600690256
36. Sheard, N. (2014). Airports and urban sectoral employment. *Journal of Urban Economics, 80*, 133-152. https://doi.org/10.1016/j.jue.2014.01.002

37. Sheard, N. (2018). *Airport size and urban growth*. *Economics of Transition and Institutional Change, 26*(2), 171-194. https://doi.org/10.1111/ecca.12262
38. Tan, Li and Bian, Shibo and Yan, Yayi and Hu, Zhiming, Generalized Impulse Response Analysis for Time-Varying VAR Models. Available at SSRN: https://ssrn.com/abstract=5243764 or http://dx.doi.org/10.2139/ssrn.5243764
39. Toda, H. Y., & Phillips, P. C. B. (1993). Vector autoregressions and causality. *Econometrica, 61*(6), 1367-1393. https://doi.org/10.2307/2951647
40. Tsiotas, D., Niavis, S., Polyzos, S., & Papageorgiou, A. (2020). Developing indicators for capturing the airports dynamics in regional and tourism development: Evidence from Greece. *Journal of Air Transport Studies, 11*(1), 31-46. https://doi.org/10.38008/jats.v11i1.153
41. U.S. Bureau of Labor Statistics. (n.d.). *Denver-Aurora-Lakewood, CO metropolitan statistical area (MSA)*. U.S. Bureau of Labor Statistics. Retrieved November 25, 2024, from https://www.bls.gov/regions/mountain-plains/co_denver_msa.htm
42. U.S. Bureau of Labor Statistics. (n.d.). *Colorado*. U.S. Bureau of Labor Statistics. Retrieved November 25, 2024, from https://www.bls.gov/regions/mountain-plains/colorado.htm
43. USTR. (2024). *Colorado State Export Data*. Office of the U.S. Trade Representative. Retrieved from https://ustr.gov/map/state-benefits/co
44. Yao, S., & Yang, X. (2008). *Airport development and regional economic growth in China* (SSRN No. 1084095). Retrieved from SSRN: https://doi.org/10.2139/ssrn.1101574

Chapter-3:
“Inter-fuel Substitution in the Aviation Sector and Sustainable Aviation Fuel: Case study of California”

3.1. Introduction

The growing need for sustainable energy to reduce the environmental impact of fossil fuels has led to increased interest in biofuels, especially for transportation. Biofuels, made from organic materials, are a renewable energy source that can help cut greenhouse gas emissions, reduce reliance on limited resources, and improve energy security. As concerns about climate change rise, the aviation industry, which is a major source of carbon emissions, is under pressure to adopt cleaner fuel alternatives. In road transport, biofuels are seen as a promising replacement for petroleum-based fuels. They can be locally produced from a variety of materials like crops, algae, and waste, making them attractive for reducing emissions and dependence on imported oil. Bioethanol and biodiesel have become widely used in internal combustion engines and have made significant progress in commercial use. However, using biofuels in aviation is more challenging because of the industry’s specific fuel requirements, such as high energy density, low freezing points, and reliable performance at high altitudes. Despite these challenges, considerable research has been done on biofuels for aviation, particularly Sustainable Aviation Fuels (SAF). SAFs are renewable, "drop-in" fuels that can replace conventional jet fuel while reducing carbon emissions. They can be blended with regular jet fuel in amounts ranging from 10% to 50%, depending on the feedstock and production method (US Department of Energy, 2021).

SAFs are made from a variety of feedstocks, including plant oils, agricultural waste, municipal solid waste, and algae. Different technologies, such as Fischer-Tropsch synthesis, hydro-processed esters and fatty acids (HEFA), and alcohol-to-jet (ATJ) processes, which are used to produce these fuels. HEFA is the most commercially advanced method, where plant oils or animal fats are hydrogenated to create a synthetic aviation fuel that meets the standards of traditional jet fuel. The Fischer-Tropsch process, on the other hand, turns biomass into synthetic fuel by gasification, producing a high-energy-density fuel suitable for aviation. SAFs can also be blended with conventional jet fuels, allowing a smooth transition to sustainable alternatives without requiring major changes to current aircraft or infrastructure (US Department of Energy, 2021).

The environmental benefits of SAFs are significant. Studies show that SAFs can cut lifecycle greenhouse gas emissions by up to 80% compared to fossil jet fuels, depending on the feedstock and production method. SAFs are also compatible with existing aircraft engines and infrastructure, offering an immediate way to reduce the aviation industry's carbon footprint. However, challenges remain, such as high production costs, limited feedstock availability, and the need for large-scale infrastructure development, which make widespread adoption difficult (IATA, 2020). In the U.S., California has become a leader in SAFs adoption due to proactive policies and programs like the Low Carbon Fuel Standard (LCFS) and the Sustainable Aviation Fuel Partnership Program. These efforts have led to significant investments in biofuel production and SAFs commercialization. While cost and infrastructure challenges persist, California's commitment to renewable energy serves as an example for other regions seeking to transition to cleaner energy in transportation and aviation (CARB, 2021).

The potential for substitution between conventional petroleum-based fuels and biofuels is crucial for successfully implementing market-based policies to reduce greenhouse gas emissions in aviation. Greater substitution possibilities between aviation fuels enhance the effectiveness of such policies. This substitution potential can be assessed through elasticity of substitution measures. Furthermore, comparing partial elasticity of substitution estimates helps evaluate the significance of alternative, market-driven policy interventions.

While there is an abundant literature about biofuels and their effect on the economy from economic, regional, and environmental angles, there is still a shortage of empirical papers that analyze the historical effect of biofuels in transportation sector in general and inter-fuel substitutability of conventional and SAFs fuels in aviation sector in particular.

Since the cost-minimizing behavior of biofuels users and California government incentives for increasing the use of biofuels jointly determine the extent to which the use of biofuels and SAFs contributes to inter-fuel substitution in the transportation and aviation sectors, the specific objective of this study is twofold. First, this study aims to examine the historical extent of substitutability or complementarity of different fuels in transportation sector in general after the introduction of biofuels from mid-2010's. Second, this study examines the extent to which the use of biofuels contributes to inter-fuel substitution particularly in the aviation sector. Estimation results from OLS and SUR models provide insights on inter-fuel substitution and biomass use in the transportation sector in general and in aviation industry in particular. This study finds evidence that, in general, the introduction of biofuels into the transportation sector does not change traditional substitute and complement relationships between different fuel types in the general transportation sector but does become a valid substitute for traditional kerosine-type jet fuel use in the aviation sector in particular in the California case study presented in this paper.

The rest of this paper is organized as follows. In Section 2 literature review about inter-fuel substitutions of biofuels and SAFs in transportation and aviation sectors and overview about California's case is presented. Section 3 presents a description of basic static and alternative time trend models with different specifications and discusses their applicability to the analysis of inter-fuel substitution. In Section 4, data descriptions are provided. In Section 5, the empirical results are provided with a focus on inter-fuel substitution in the transportation sector of California in general and in the aviation industry in particular. Finally, conclusions and policy implications are provided in Section 6.

3.2. Literature Review

3.2.1. Inter-Fuel Substitutions of Biofuels in the Transportation Sector

The transportation sector uses a lot of energy and produces a large share of global greenhouse gas emissions, making it an important area for biofuels to replace traditional fossil fuels. Research on fuel alternatives shows how biofuels like ethanol and biodiesel can replace regular fuels in transportation, both in the U.S. and around the world. These studies highlight the potential of biofuels to cut emissions and offer a more sustainable option for transportation, helping reduce dependence on fossil fuels and contribute to more sustainable climate change.

Price Sensitivity Impact: Studies on price sensitivity show that the prices of biofuels and fossil fuels, influenced by subsidies and regulations, are key factors in deciding how much one fuel can replace another. Hochman and Timilsina (2017) found that price increases have a limited effect on reducing CO_2 emissions because it's hard to switch between different fuels. They concluded that policies encouraging fuel-efficient vehicles are more effective than simply changing fuel prices. Anderson (2012) looked at how household demand for ethanol as a substitute for gasoline responds to price changes, finding that a $0.10 price increase in ethanol led to a 12-16% drop in demand. Xie and Hawkes (2015) studied fuel substitution in China's transport sector and found greater potential for switching between oil and natural gas compared to other energy sources. Edelenbosch et al. (2017) simulated the impact of fuel price shocks, predicting that a doubling of oil prices could reduce oil's market share by 50% by 2050, with biofuels and electricity filling the gap.

Suh (2019) focused on the effects of biofuels on CO_2 emissions in the U.S. transportation sector, finding that while petroleum demand is less responsive to price changes, ethanol can reduce petroleum use, especially when prices rise. Uchôa et al. (2019) looked at the demand for gasoline, ethanol, and diesel in Brazil, finding that price changes affected demand, but GDP growth had a stronger effect on gasoline and diesel demand. Hossain and Serletis (2020) found small but significant substitution between biofuels and fossil fuels in the U.S. from 1990 to 2017. Brito et al. (2020) examined how fuel price changes influenced vehicle market shares in Brazil, noting that flex-fuel vehicle adoption was more influenced by operating costs than price changes. Finally, Neves and Marques (2021) found that while alternative energy sources in U.S. transportation reduced CO_2 emissions, they also limited economic growth, which in turn affected emissions.

Taxes, Subsidies, Mandates: Both U.S. and international studies highlight how government policies, like taxes, subsidies, and mandates, influence the adoption of biofuels, especially ethanol blends. Kim et al. (2011) studied the impact of a carbon tax in Korea's transportation sector, finding that such a tax could reduce CO_2 emissions by up to 1,090,325 tons if more efficient vehicles were available. Bartocci and Pisani (2013) looked at the effects of taxing motor vehicle fuels in France, Germany, Italy, and Spain, finding that reducing electricity consumption taxes and increasing renewable electricity subsidies could lower CO_2 emissions in the transport sector and encourage renewable energy use. Wu and Langpap (2014) used a model to analyze the effects of biofuel mandates and subsidies in the U.S., showing that while biofuels increased food prices and decreased gasoline prices, they had little overall impact on consumer welfare. Shriver (2015) examined how the availability of ethanol-compatible vehicles and ethanol fuel affected demand, finding that more ethanol retailers led to a 6% increase in ethanol vehicle purchases. Finally, Bardazzi et al. (2024) studied gasoline and diesel demand in Italy and found that household demand for transport fuels was more responsive to tax changes than to oil price fluctuations, suggesting that lowering fuel taxes could help achieve decarbonization goals in transportation.

Infrastructure and Technology: The availability of biofuel-compatible vehicles (like flex-fuel vehicles or FFVs) and refueling stations plays a big role in how often people switch to alternative fuels. Huse (2018) studied 12 years of data from Sweden to understand fuel choices made by drivers of multi-fuel vehicles (MFVs), specifically between petrol and ethanol. His results showed that most drivers choose petrol when its price is similar to ethanol's. Huse's policy simulations also showed that fossil fuel taxes aren't very effective in encouraging people to switch to alternative fuels and may cause economic problems. The study emphasizes the need to consider how fuel choices are made when creating policies to promote MFVs and alternative fuels.

Environmental Sustainability: The environmental impact of biofuel adoption is crucial when determining their effectiveness as alternatives. Chen et al. (2014) created a model to study the effects of three policies: the Renewable Fuel Standard (RFS), Low Carbon Fuel Standard (LCFS), and a carbon tax, focusing on welfare and greenhouse gas (GHG) emissions. Their simulations, which adjusted policies to achieve the same U.S. GHG emissions, showed that while the RFS cuts fossil fuel use by promoting food-crop-based biofuels, it raises food prices and leads to a smaller global GHG reduction compared to the LCFS and carbon tax. All three policies improved social welfare in the U.S., but the carbon tax was the most efficient worldwide. Nuñez and Önal (2016) studied the impact of taxes, subsidies, and blending mandates on Brazil's ethanol and sugar markets using an economic simulation model. They found that low ethanol blending rates decreased demand for conventional vehicles but had little effect on flex-fuel and ethanol-only vehicles. Lowering fuel taxes made sugarcane ethanol less competitive, reducing producer profits, while improving social welfare but increasing GHG emissions. Nugroho and Zhu (2024) examined the planning of agricultural biomass supply and transportation in hydrogen and syngas supply chains, using a bi-level optimization model. Their findings showed how suppliers could optimize prices, production capacity, and order distribution between biomass suppliers and biofuel producers, balancing profits and minimizing supply chain costs while enhancing the substitution potential of different fuels and biomass types.

3.2.2. Inter-Fuel Substitution of Biofuels in the Aviation Sector

The aviation sector is a challenging area for decarbonization due to its heavy reliance on fossil-based jet fuels, high energy-density requirements, and limited technological alternatives. Biofuels, particularly SAFs, are gaining attention as viable substitutes. This section reviews key empirical studies focused on inter-fuel substitution of biofuels in the aviation sector, highlighting their methodologies and findings in both U.S. and international contexts.

Cost and Policy Support: Studies consistently show that SAFs are much more expensive than traditional fossil-based jet fuels. To address this cost gap, government incentives such as subsidies, quotas, and carbon pricing are essential. Membreno (2024) used econometric and machine learning models to examine how factors like GDP, oil prices, population, and seasonality affect aviation travel demand. The study found that business passengers are less sensitive to price changes, while economy passengers exhibited little to no price sensitivity, with some routes even showing a positive price elasticity, challenging the typical assumption that economy passengers are more price-sensitive. Chao et al. (2019) combined airline operations models with multi-feedstock SAFs and life-cycle assessments to evaluate policies like CORSIA. Their simulations suggested that such policies could reduce air travel growth due to higher fares, but by 2050, U.S. airline GHG emissions could be reduced by 37.5–50%, depending on factors like fuel prices and carbon pricing. Winchester et al. (2013) assessed the cost of renewable jet fuel production using the HEFA process and found that meeting biofuel targets by 2020 would require substantial subsidies. While aviation biofuel consumption had limited impact on jet fuel prices and CO_2 emissions, it did affect with ground transportation fuel mandates. Reimer and Zheng (2017) used a general equilibrium model to assess the feasibility of a camelina oilseed-based biofuel supply chain in the Pacific Northwest, U.S., concluding that the supply chain could be viable if passengers were willing to pay more for biofuels. Policy interventions, like subsidies and taxes on fuels, could support this supply chain. Mueller (2024) analyzed the economic impact of Power-to-Liquid (PtL) fuel blending quotas and price policies in Germany. The study showed that low blending quotas mainly affected the aviation sector by increasing consumer prices and reducing aviation output, while higher quotas led to macroeconomic losses, including a slight decline in GDP. However, PtL production subsidies helped ease some of these impacts.

Feedstock Challenges: The availability, cost, and competition for land and resources are key concerns when it comes to adopting SAFs. McCollum et al. (2021) looked at the potential supply of oilseed crops, like canola and rapeseed, for SAFs production. They surveyed U.S. farmers in the Great Plains and Pacific Northwest to understand their willingness to grow these crops under contract. The study found that Kansas and North Dakota would have the largest areas dedicated to oilseed crops, and crop prices were very sensitive to supply. The research showed that offering favorable contract terms and crop insurance could help boost supply without needing significant price increases. Zhao et al. (2021) estimated the emissions from land use changes (ILUC) for 17 SAFs production pathways using an economic model. They found that the emissions varied greatly, with one pathway (USA miscanthus alcohol-to-jet) showing negative emissions (-58.5 g $CO_2e\ MJ^{-1}$) due to carbon sequestration. Most of the pathways examined had lower emissions over their entire life cycle than traditional jet fuel, highlighting the potential of SAFs, especially those made from cellulosic materials, to significantly reduce emissions from aviation.

3.2.3. Biofuels and Sustainable Aviation Fuels in California

California has been a leader in renewable energy, especially in biofuels and SAFs. The state has passed strong laws, offered financial support, and worked with businesses and universities to promote biofuels and SAFs in transportation and aviation. These efforts are part of California's goal to become carbon neutral by 2045.

In transportation, California has pushed for biofuels like ethanol and biodiesel as cleaner alternatives to fossil fuels. The Low Carbon Fuel Standard (LCFS), created by the California Air Resources Board (CARB) in 2009, aims to cut the carbon intensity of transportation fuels by 20% by 2030. Biofuels are key to reaching this goal because they produce fewer carbon emissions than regular fossil fuels. The LCFS uses a credit system where biofuels, like ethanol and biodiesel, earn tradable credits for being low in carbon. This encourages the use of biofuels made from things like agricultural waste, algae, and used oils, helping to reduce the need for petroleum-based fuels.

California has been a leader in developing advanced biofuels made from non-food materials like waste oils, garbage, and algae. This helps avoid the problem of using food crops like corn and soybeans for fuel, which can raise concerns about food supply. Programs such as the Advanced Biofuel Feedstock Incentive give financial support to biofuel producers and help build the infrastructure needed, encouraging new ideas in biofuel production (CARB, 2022).

In aviation, California has made strong progress in promoting SAFs. SAFs are made from renewable sources like waste oils, algae, and leftover crops, offering a way to reduce the carbon footprint of flying. SAFs are a "drop-in" replacement for traditional jet fuel, meaning they can be used in current airplanes and systems with little change. This makes SAFs a practical and immediate option for reducing emissions, with the potential to cut greenhouse gases by up to 80%, depending on the materials used to make them (IATA, 2020).

To support the development of SAFs, California has introduced several important programs. In 2018, the California Air Resources Board (CARB) updated the Low Carbon Fuel Standard (LCFS) to include alternative jet fuels as an "opt-in" option. While traditional jet fuel doesn't earn credits, SAFs can receive credits based on how much they reduce carbon emissions over their lifecycle (Berkley Law, The Center for Law, Energy & the Environment, 2022). California also set a carbon intensity benchmark for fossil jet fuel at 89.37 gCO_2e per megajoule, which will gradually decrease in line with the state's climate targets (Fig.-3.1). Starting in 2023, the SAFs benchmark will align with the diesel fuel benchmark, meaning SAFs will earn credits just like diesel substitutes if their carbon emissions are similar (Berkley Law, The Center for Law, Energy & the Environment, 2022). Additionally, the California Sustainable Aviation Fuel Partnership, a collaboration between CARB and Airlines for America (A4A), aims to make 200 million gallons of SAFs available in California by 2035, meeting about 40% of the state's aviation fuel needs (CARB, 2024). With its strong policies, growing low-carbon fuel infrastructure, and high aviation emissions, California is in a great position to lead the U.S. in promoting cleaner aviation fuel.

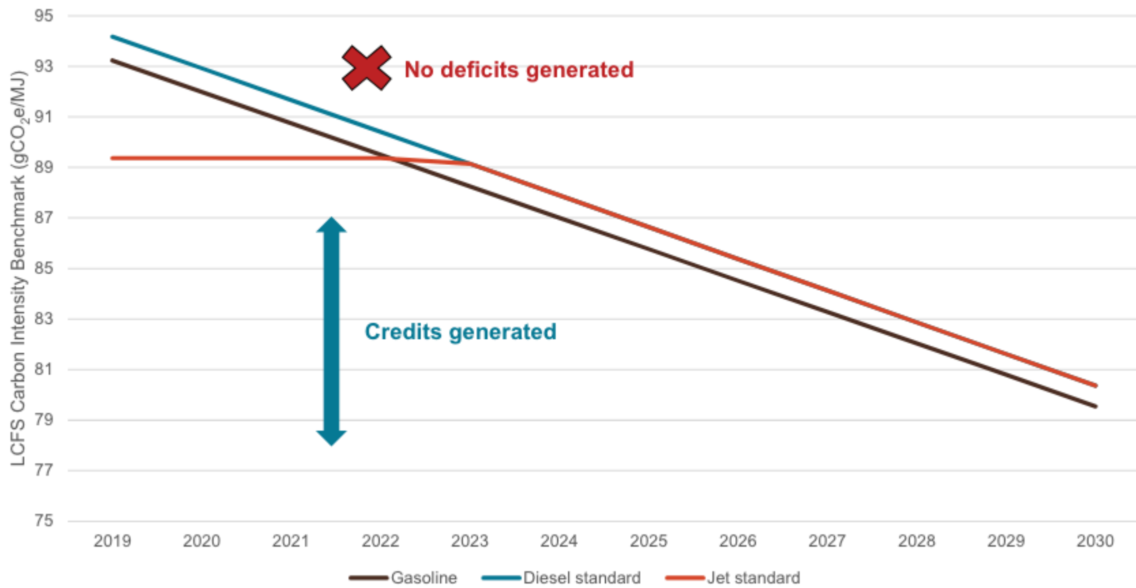


Figure-3.1. Comparison of benchmark carbon intensities for gasoline, diesel, and jet fuels in the California Low Carbon Fuel Standard, 2019-2030

Source: *International Council on Clean Transportation (ICCT), 2022.*

The price of Sustainable Aviation Fuel (SAF) is closely connected to government policies that support cleaner fuels. The California's Low Carbon Fuel Standard (LCFS) offer credits to companies that produce or use low-carbon fuels like SAF[5]. These credits have their own market prices, which can go up or down depending on supply, demand, and policy changes. When credit prices are high, SAF producers can earn more money, which helps lower the effective cost of SAF.

---

[5] California Air Resources Board (CARB) – LCFS Program: https://ww2.arb.ca.gov/our-work/programs/low-carbon-fuel-standard. CARB provides data on LCFS credit prices and explains how the policy incentivizes low-carbon fuels.

But if credit prices fall, SAF becomes more expensive and less competitive compared to regular jet fuel. This shows that SAF prices don't just depend on production costs—they are also shaped by the value of these policy credits. By looking at the data on credit prices, we can see how much SAF prices are affected by biofuel policies and how important these programs are in making cleaner aviation fuel more affordable.

3.3. Model

3.3.1. Basic Static Model

Based on the previous literature (Basso et al. 2007, Lin et al. 2013), the fuel demand D can be expressed as a function of fuel prices P, and other determinants X of fuel demand. This model can be written as:

$$D = f\ (P, X) \qquad (3.1)$$

The equation (3.1) is estimated using simple "static" reduced-form demand model, where demand for fuel i is a function of prices of fuel i and j. The double log model is used, as it is found to be more appropriate for fuel consumption than linear model, based on previous literature (Dahl 1986, Espey 1998).

In the basic double-log model, it is assumed that the elasticity remains constant throughout the entire analysis period.:

$$lnD_{it} = \beta_0 + \beta_1 lnP_{it} + \beta_2 lnP_{jt} + \varepsilon_{it} \qquad (3.2)$$

$$lnD_{jt} = \beta_0 + \beta_1 lnP_{it} + \beta_2 lnP_{jt} + \varepsilon_{jt} \qquad (3.3)$$

where D_{it} is demand for fuel i in gallons at time t, D_{jt} is demand for fuel j in gallons at time t, P_{it} is the price of fuel i at time t, P_{jt} is the price of fuel j at time t, and ε_t is a mean zero error term.

The own-price elasticity is expected to be:

$$\frac{\partial lnD_{it}}{\partial lnP_{it}} = \beta_1 \qquad (3.4)$$

The cross-price elasticity is expected to be:

$$\frac{\partial lnD_{it}}{\partial lnP_{jt}} = \beta_2 \qquad (3.5)$$

Under the simple assumption about the coefficients β_1 and β_2 they are expected to represent the estimates of the price elasticity of demand when the observed price and quantity are in equilibrium (Lin et al., 2013).

The coefficients in this static model may be difficult to interpret due to omitted variable bias. One key variable that could be left out is time. This static model does not account for the time effect, meaning it does not consider how demand changes over time in response to price or other factors. However, in reality, there might be delays in reaching equilibrium due to various reasons, such as imperfections in alternative fuel markets. As a result, the elasticity estimates from the static model may only reflect current adjustments and might differ when time-related control variables are included.

3.3.2. Analyzing the impact of time trends using alternative specification

When working with time series data, it's important to carefully consider co-movements, as they may indicate trending behavior (Wooldridge, 2008). It's crucial to check whether significant time trends exist in the data. If a shared trend is overlooked, the analysis might lead to a spurious regression, where both the dependent and independent variables appear to be correlated, not because of a true relationship, but due to the influence of the omitted time trend.

To check if the data is exhibiting constant increments over time, we include an exponential trend and an exponential trend squared into the basic static model (Wooldridge, 2008). The new updated model specification for two different fuels is:

$$lnD_{it} = \beta_0 + \beta_1 lnP_{it} + \beta_2 lnP_{jt} + \beta_3 Trend_{it} + \beta_4 {Trend_{it}}^2 + \varepsilon_{it} \quad (3.6)$$

$$lnD_{jt} = \beta_0 + \beta_1 lnP_{it} + \beta_2 lnP_{jt} + \beta_3 Trend_{jt} + \beta_4 {Trend_{jt}}^2 + \varepsilon_{jt} \quad (3.7)$$

where, fuel i is one type of fuel and fuel j is another type of fuel.

In particular, we hypothesize that if a time trend is necessary, it will appear as a significant factor influencing the dependent variable. The time trend will capture the general upward or downward movement over time, while the time trend squared will indicate the rate of change. The coefficients of the independent variables will reflect their relationship with the dependent variable, holding other factors constant and accounting for the trend in the dependent variable (Wooldridge, 2008). We will estimate the model using standard OLS regression in STATA-18 (2024).

3.3.3. Checking the Model Accuracy Using Alternative Regression: Seemingly Unrelated Regression

Since the model specification that we use above has the same time trend variables for both types of fuel (so, same independent variables), it is possible to expect that the equation errors in both equations might be correlated. To deal with such a model, where a set of equations might be correlated through their error terms, SUR is widely used (Zellner, 1963).

The fuel allocation model estimates the equations (3.6, 3.7) above for fuel (i) and fuel (j) simultaneously by iterative seemingly unrelated regression (ITSUR).

The SUR model permits nonzero covariance between the error terms ε_{it} and ε_{jt} for a given fuels (i) and (j) across both equations (3.6, 3.7) and independent variables: $Cov(\varepsilon_{it}, \varepsilon_{jt}) = \sigma_{ijt}$, while assuming: $Cov(\varepsilon_{it}, \varepsilon_{jt}) = 0$ if $i \neq j$.

It is the potential nonzero covariance across both fuel (i) and fuel (j) equations (3.6, 3.7) that allows for an improvement in efficiency of SUR relative to the classical OLS estimator of each $\beta's$ (Zellner, 1963).

The SUR model estimates the correlations between the equations and tests for error independence using the Breusch-Pagan test, with the null hypothesis that the error terms in the equations are *correlated* (Zellner, 1963). The covariance matrix is estimated iteratively from the residuals generated by the initial parameter estimates. This estimated covariance matrix is updated until the parameter estimates converge (Cherubini, 2011).

3.4. Data

Since the paper consists of two parts, two different datasets are used in each part.

3.4.1. Data on Primary Transportation Fuels

This dataset is mainly used to reveal historical relationships between fuel usage in aviation and other transportation sectors (cars, ships, trains, buses, trucks, etc.), and particularly after introduction of biofuels from early 2010's. The annual data is obtained from the Annual Energy Review published by the U.S. Energy Information Administration (EIA) particularly for California. The data represents the annual quantities and prices of primary fuels used in the transportation sector in California for the period from 1970 to 2021: petroleum products with included biomass-based fuels and electricity. Since the obtained price data from EIA was in nominal values, the EIA Real Prices Convertor spreadsheet is used to obtain real prices, with a base year being November 2024 (Ref: https://www.eia.gov/outlooks/steo/realprices/).

The data for this analysis is interpreted as that resulting from General Equilibrium processes (rather than Partial Equilibrium) as it shows how supply and demand interact and tend toward a balance in an economy of multiple markets working at once. This data is recorded after suppliers and consumers have already made their independent decisions (EIA).

This analysis includes different categories of fuels. Based on the categories of the EIA, the types and uses of transportation energy sources include:

- Motor gasoline used in cars, motorcycles, light trucks, and boats.
- Aviation gasoline usually used in small airplanes.
- Distillate fuels used mainly by trucks, buses, trains and in boats and ships.
- Jet fuel used in jet airplanes (large) and some types of helicopters.
- Residual fuel oil used in ships.
- Biofuels blended into gasoline and diesel fuel.
- Propane (a hydrocarbon gas liquid) (HGL) used in cars, buses, and trucks (often, vehicles in government and private vehicle fleets).
- Electricity used by public mass transit systems and by electric vehicles.

According to EIA, biomass-based biofuels are not defined as a separate category in this dataset, but instead are included into the following series:

- Beginning in 1993, includes fuel ethanol blended into motor gasoline (EIA).
- Beginning in 2009, biodiesel blended into distillate fuel oil (EIA).
- Beginning in 2011, renewable diesel blended into distillate fuel oil (EIA).

- Beginning in 2021, includes biodiesel and renewable diesel product supplied into distillate fuel oil (EIA).

Table-3.1 shows all fuel quantities consumed by the transportation sector, measured in trillions of Btu. (*British thermal unit*), and their prices, expressed in dollars per million Btu.

Table-3.1. Summary Statistics: Primary Fuels Data in Transportation Sector in California, 1970-2021

Variable	Obs.	Mean	Std.dev.	Min	Max
Quantities (trillion Btu.)					
Aviation Gasoline	52	4.62	2.85	0.7	11
Distillate Fuel Oil	52	347.92	100.45	171.5	494.3
HGL (Propane)	52	2.27	1.36	0.4	6.6
Jet Fuel	52	469.02	101.24	316.7	628.2
Motor Gasoline	52	1581.08	231.91	1106.5	1956.8
Residual Fuel Oil	52	213.74	76.87	84.5	419.2
Electricity	52	1.68	0.91	0.2	3.1
Prices ($/million BTU.)					
Aviation Gasoline	52	26.00	9.21	12.02	45.37
Distillate Fuel Oil	52	23.27	8.80	7.86	42.31
HGL (Propane)	52	23.61	7.55	7.48	41.70
Jet Fuel	52	15.38	7.48	4.04	32.57
Motor Gasoline	52	27.30	8.22	15.45	44.27
Residual Fuel Oil	52	12.40	7.43	1.99	31.97
Electricity	52	31.94	10.33	15.95	60.03

Source: The U.S. Energy Information Administration (EIA).

According to the data, the transportation sector is the most dependent on the use of motor gasoline, jet fuel, distillate fuel oil and residual fuel oil, and is the least dependent on the use of aviation gasoline, propane and electricity (Fig.-3.2). The reason for this is because most refineries in the U.S. focus on producing motor gasoline, jet fuel and distillate fuel (diesel), because of higher demand/profitability. Refineries produce other types of fuel based on volatile market demand. One reason why aviation gasoline volume is so much less compared to other aviation fuel is because aviation gasoline has a lead-based formula, which makes it more dangerous to human health than jet fuel (EPA, 2022). Also, the Federal Aviation Administration (FAA), which funds and sets regulations for aircraft and airports around the country, announced in 2022 its plans to phase out the fuel by 2030 (Ref: https://collective.coloradotrust.org/stories/in-communities-near-some-colorado-airports-concerns-persist-over-leaded-aviation-fuel/).

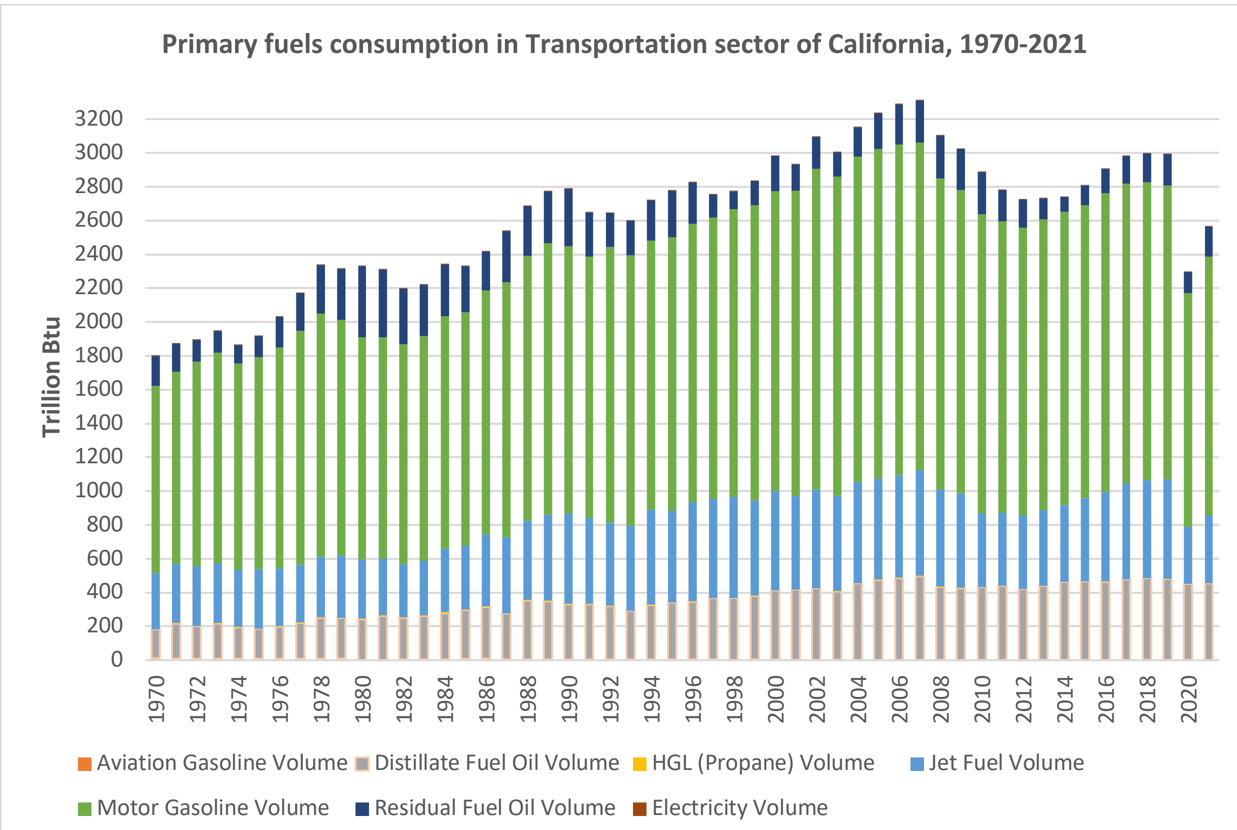


Figure-3.2. Energy consumption in the Transportation Sector in California by source, 1970-2021

Source: The U.S. Energy Information Administration (EIA).

Also, the price data by energy sources in transportation sector of California (Fig.-3.3) reveals that electricity, residual fuel oi, motor gasoline, jet fuel and propane demonstrate volatile activity, with huge increases for electricity, residual oil and motor gasoline over time, while the price for distillate fuel oil and aviation gasoline are more stable.

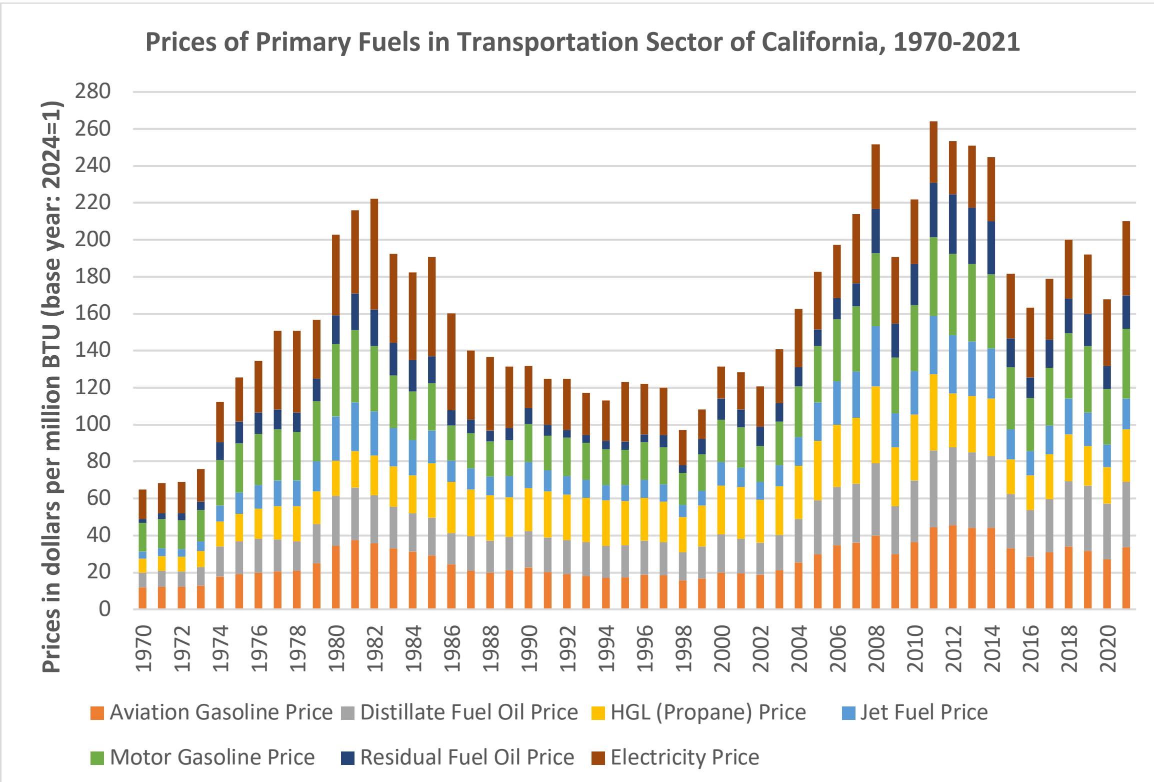


Figure-3.3. Energy prices in the California Transportation Sector by source, 1970-2021

Source: The U.S. Energy Information Administration (EIA).

3.4.2. Data on Primary Aviation Fuels

This aviation fuels dataset is mainly used to reveal historical relationships particularly in the aviation sector: if there is any competition between jet fuel and sustainable aviation fuel. The monthly data for California is obtained from the U.S. Energy Information Administration (EIA) and from the EPA. The data represents the monthly quantities and prices of primary aviation fuels used in the aviation transportation sector in California for the period from 2013 to 2022: jet fuel and SAFs.

The monthly data is obtained from two sources:

- EIA: California kerosene-type monthly Jet Fuel volumes and prices series (January 2013-March 2022)
- EPA: California Renewable Fuel Standard program data (January 2013-December 2022):
    - Aggregated monthly data on renewable fuel volume production (fuel categories - D.4, D.5)
    - Weekly volume weighted average RIN prices (fuel categories - D.4, D.5)

Since the obtained price data from EIA and EPA was in nominal values, the EIA Real Prices Convertor spreadsheet is used to obtain real prices, with a base year being November 2024 (Ref: https://www.eia.gov/outlooks/steo/realprices/).

The data for this analysis also can be seen as being the result of a General Equilibrium process for the same reasons as with the previous dataset. According to EPA (see 40 CFR 80.1426.), fuel categories D.4 and D.5 mostly include biodiesel, renewable diesel, jet fuel, and heating oil fuels – so, they can serve as *a proxy of sustainable aviation fuel* in those categories. Also, even though the SAFs data comes from EPA and reflects the US national volumes and prices traded, but according to EIA (Ref: https://www.eia.gov/todayinenergy/detail.php?id=57180), almost all proxies of SAFs in the USA is consumed in California. That's why this national data from EPA could serve as a proxy for SAFs traded volumes and prices in California. Weekly prices are converted to monthly prices by picking up the highest price on one date during the month (middle of month: 15-th day).

Table-3.2 shows all fuel quantities consumed by the aviation sector, measured in gallons per day, and their prices, expressed in dollars per gallon.

Table-3.2. Summary Statistics: Primary Fuels Data in Aviation Sector in California, 2013-2022

Variables	Obs.	Mean	Std.dev.	Min	Max
Quantities (gallons per day)					
Jet Fuel Volume	111	11,926,796.42	1,090,885.93	4,067,600	14,215,800
SAF Volume	111	446,727	495,990	41,590	2,922,508
Prices ($ per gallon)					
Jet Fuel Price	111	2.52	0.82	0.90	4.41
SAF Price	111	1.08	0.38	0.53	2.10

Source: The U.S. Environmental Protection Agency (EPA) and The U.S. Energy Information Administration (EIA).

According to the data, the air transportation sector in California is the most dependent on the use of traditional kerosine-based jet fuel, which constitutes 98% of consumption, but the least on the use of biomass-based SAFs, which constitutes just 2% of total consumption (Fig.-3.4).

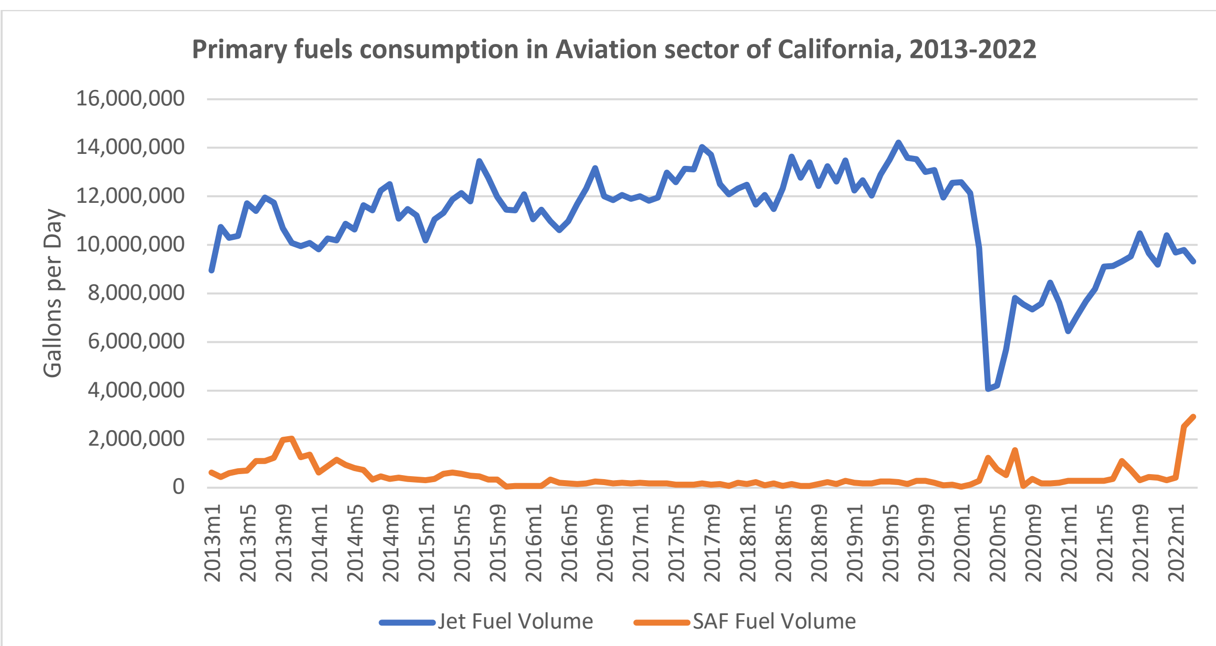


Figure-3.4. Energy consumption in the California Air Transportation Sector by source, 2013-2022

Source: The U.S. Environmental Protection Agency (EPA) and The U.S. Energy Information Administration (EIA).

Also, the price data by energy sources in air transportation sector of California (Fig.-3.5) reveals that traditional kerosine-based jet fuel and SAF demonstrated opposite movement till 2020's, and later experienced increasing trend in both types of fuel. It is worth noting that, even though it is expected that SAF should be priced with a higher price by theory, the data shows that actually it is cheaper than jet fuel throughout the analyzed data period. The reason for this is because the price data is just *a proxy for SAF* (not actual SAF prices), since it is a price of combination of different renewable biomass-based fuels in D.4 and D.5 categories of the U.S. Environmental Protection Agency (EPA) that include: biodiesel, renewable diesel, renewable jet fuel, and heating oil, naphtha, ethanol (Ref: https://www.law.cornell.edu/cfr/text/40/80.1426). (Ref: see 40 CFR 80.1426. https://www.ecfr.gov/current/title-40/chapter-I/subchapter-C/part-80/subpart-M/section-80.1426). This fact is considered as *a limitation* of the present study.

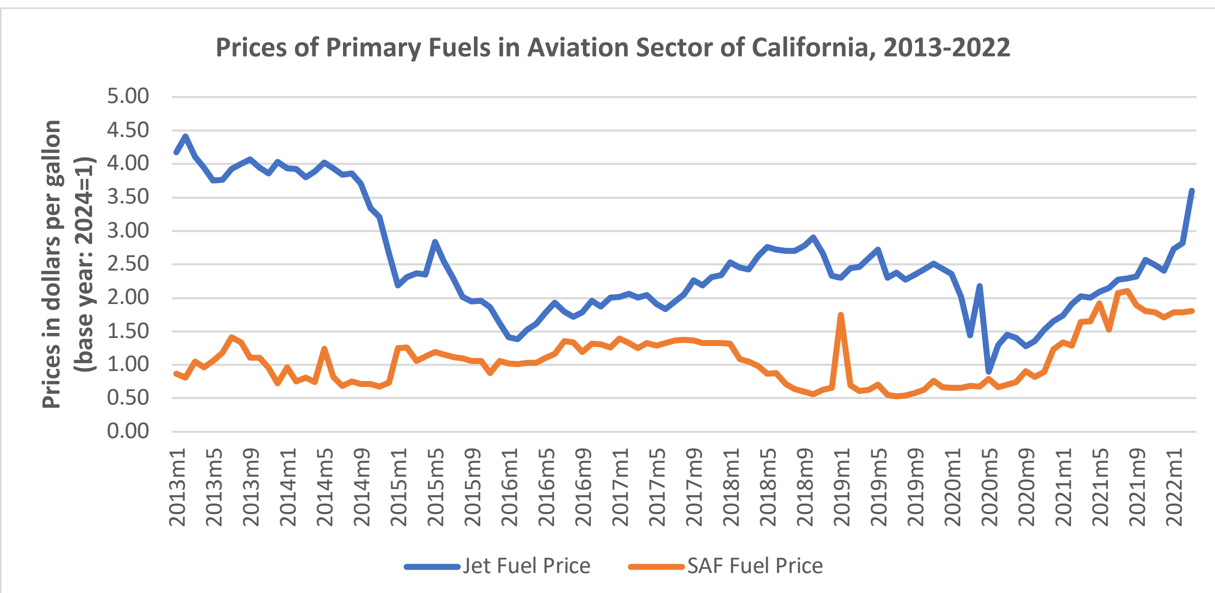


Figure-3.5. Energy prices in the California Air Transportation Sector by source, 2013-2022

Source: The U.S. Environmental Protection Agency (EPA) and The U.S. Energy Information Administration (EIA).

Another *limitation* of the current study is that only the *spot prices* for SAF and jet fuel are used in the analysis. Theoretically, it would be more preferred to use *futures prices* for both types of aviation fuels, because airlines might prefer using financial products like futures prices to smooth the price volatilities of aviation fuels market. But unfortunately, as of June, 2025, the monthly futures prices data for aviation fuels in California (jet fuel and SAF) is not available[6]. This could be a possible avenue for the future work as a robustness check, when data becomes available.

3.5. Results

3.5.1. Results on Different Fuels Substitutability in Transportation sector in California

In this first part of the analysis, the historical relationships between fuel usage in aviation and other transportation sectors (cars, ships, trains, buses, trucks, etc.) are presented. The analysis is done based on 1970 – 2021 annual data from EIA about primary fuels used (volumes sold and selling prices) in the transportation sector in California. The OLS with time trend specification is used in all regressions. Regressions vary by dependent variables, representing different types of fuels used in different areas of transportation sector (EIA):

- Aviation gasoline - small airplanes.
- Propane (a hydrocarbon gas liquid - HGL) - cars, buses, and trucks.
- Jet fuel - jet airplanes (large) and some types of helicopters.
- Motor gasoline - cars, motorcycles, light trucks, and boats.

[6] Alternative spot prices for jet fuels in California could be found at: https://www.opis.com/product/pricing/spot/west-coast-spot-market-report/. But monthly futures prices data for jet fuel and SAF is not available to date as of authors' knowledge.

- Electricity - public mass transit systems, electric vehicles.
- Distillate fuels - trucks, buses, trains, boats and ships.
- Biofuels may be blended into gasoline, diesel fuel, and jet fuel.
- Residual fuel oil - ships.

Note: Regression results for Distillate fuels and Residual fuel are excluded from the paper due to:

- Distillate fuels– regression results reveal that none of variables are significant when distillate fuel is considered as dependent variable, that's why the regression results are not included into the paper, but are available in the coding file.
- Residual fuel - this fuel type may have an inferior good property, so it is different and difficult to interpret compared to other fuel types, that is why the regression results are not included into the paper.

The results of Model-7 for each fuel type with all other fuels regressing simultaneously ("sink type" regression) are presented in each of the Tables 3.3-3.7 for reference only, due to high correlation between variables. The correlation matrix of all variables is presented in the Appendix-C (Table-C1).

3.5.1.1. Aviation Gasoline

The results of regression with *Aviation Gasoline* as a dependent variable are presented in Table-3.3.

Table-3.3. Substitutability between Dependent variable: *Ln_Aviation Gasoline Volume* and other fuels in General Transportation sector in California in 1970-2021

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Ln _Aviation Gasoline Price	-0.698***						-1.928**
	(0.182)						(0.898)
Ln_Distillate Fuel Price		-0.763***					0.261
		(0.205)					(0.809)
Ln_HGL Price			-0.448				0.225
			(0.303)				(0.46)
Ln_Jet Fuel Price				-0.426***			1.831**
				(0.119)			(0.867)
Ln_Motor Gasoline Price					-0.909***		-2.581***
					(0.204)		(0.88)
Ln_Electricity Price						-0.305*	0.199
						(0.177)	(0.238)
Time Trend	-0.0392***	-0.0295*	-0.0154	-0.0336**	-0.0543***	-0.0445***	-0.127***
	(0.0143)	(0.0150)	(0.0265)	(0.0148)	(0.0137)	(0.0157)	(0.033)
Time Trend Squared	0.000313	0.000233	-0.000181	0.000157	0.000592**	0.000245	0.00204***
	(0.000259)	(0.000261)	(0.000421)	(0.000264)	(0.000259)	(0.000287)	(0.00056)
Constant	4.315***	4.247***	3.301***	3.192***	5.191***	3.325***	10.43***
	(0.542)	(0.542)	(0.683)	(0.292)	(0.661)	(0.608)	(2.733)
Observations	52	52	52	52	52	52	52
R-squared	0.684	0.679	0.604	0.674	0.707	0.610	0.766

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Authors calculations.

According to the results, the own-price elasticity of aviation gasoline is negative and complies with a theory: as the price of aviation fuel (small airplanes) goes up, its quantity consumed is decreasing. All other fuels used in transportation sector, beside HGL (propane in cars/buses) , like distillate fuel (diesel in cars/buses/ trains/ship), jet fuel (big airplanes), motor gasoline (cars, motorcycles, light trucks, and boats), electricity (electric vehicles), demonstrate complementary relationship with aviation gasoline when regressed separately (Models 1-6): as their prices increases by 1% the aviation gasoline volume consumed decreases with a varying magnitudes in the range of 0.3-0.9%. So, the distillate fuel, jet fuel and motor gasoline seem to be stronger complements with higher significance to aviation gasoline than electricity, likely reflecting the primary focus of refineries to produce highly-demanded aviation/jet fuels, distillate (diesel) fuel and motor fuel.

In addition, the results show that the demand for aviation gasoline is very inelastic with respect to the own-price and cross-price changes in distillate fuel, jet fuel, motor gasoline and electricity: quantity of aviation gasoline changes slower than price changes in above-mentioned fuels. This result suggests that aviation gasoline historically plays a complementary role in the fuels market, since it is used mainly in small-size aviation. So, when prices for all other transportation fuels rise, people prefer to fly less on small/private jets, probably because of higher/more expensive costs for airplane tickets. Also, the time trend analysis across all models reveals a gradual increase in aviation gasoline usage in majority of models, indicating that its adoption continues to grow.

3.5.1.2. HGL (Propane)

The results of regression with *HGL (propane)* as a dependent variable are presented in Table-3.4.

Table-3.4. Substitutability between Dependent variable: *Ln_ HGL (propane) Volume* and other fuels in General Transportation sector in California in 1970-2021

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	Model-1	Model-2	Model-3	Mode- 4	Model-5	Model-6	Model-7

Ln _Aviation Gasoline Price	0.402*						-1.009
	(0.229)						(1.154)
Ln_Distillate Fuel Price		0.287					-0.571
		(0.262)					(1.039)
Ln_HGL Price			0.326				-0.982
			(0.350)				(0.591)
Ln_Jet Fuel Price				0.232			2.339**
				(0.149)			(1.114)
Ln_Motor Gasoline Price					0.235		-2.215*
					(0.274)		(1.131)
Ln_Electricity Price						0.568***	0.692**
						(0.187)	(0.306)
Time Trend	0.0689***	0.0667***	0.0504	0.0660***	0.0752***	0.0688***	0.0699
	(0.0180)	(0.0191)	(0.0306)	(0.0186)	(0.0184)	(0.0170)	(0.0424)
Time Trend Squared	-0.00178***	-0.00174***	-0.00143***	-0.00170***	-0.00184***	-0.00171***	-0.00125*
	(0.000327)	(0.000333)	(0.000486)	(0.000331)	(0.000348)	(0.000310)	(0.00072)
Constant	-0.832	-0.407	-0.395	-0.158	-0.428	-1.548**	6.721*
	(0.685)	(0.693)	(0.789)	(0.366)	(0.890)	(0.657)	(3.512)
Observations	52	52	52	52	52	52	52
R-squared	0.523	0.505	0.501	0.517	0.500	0.571	0.636

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Source: Authors calculations.

Results suggest that aviation fuel (small airplanes) and electricity demonstrate a substitution relationship with HGL (propane) when regressed separately (Models 1-6): as their prices increase by 1% the HGL (propane) volume consumed increases with a varying magnitude of 0.4%, 0.5%, accordingly.

The findings also indicate that the demand for hydrocarbon gas liquids (HGL), particularly propane, is highly inelastic to cross-price fluctuations in aviation gasoline and electricity. This means that changes in the price of these fuels have only a minimal effect on the quantity of HGL (propane) consumed. Historically, HGL (propane) has acted as a competitive substitute in the fuel market, particularly against aviation and electricity, as it is predominantly used in vehicles like cars and buses. Consequently, when the cost of small airplanes or electric vehicles usage increases, consumers tend to shift toward using vehicles powered by propane. Additionally, the time trend analysis across all models reveals a gradual increase in HGL (propane) usage, albeit at a decelerating rate, indicating that while its adoption continues to grow, the pace of growth is slowing over time.

3.5.1.3. Jet Fuel

The results of regression with *Jet Fuel* as a dependent variable are presented in Table-3.5.

Table-3.5. Substitutability between Dependent variable: *Ln_Jet Fuel Volume* and other fuels in General Transportation sector in California in 1970-2021

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	Model-1	Model-2	Model-3	Mode- 4	Model-5	Model-6	Model-7
Ln _Aviation Gasoline Price	-0.285***						-0.690***
	(0.0501)						(0.252)
Ln_Distillate Fuel Price		-0.301***					-0.306
		(0.0584)					(0.227)
Ln_HGL Price			-0.329***				-0.144
			(0.0841)				(0.129)
Ln_Jet Fuel Price				-0.166***			0.703***
				(0.0342)			(0.243)
Ln_Motor Gasoline Price					-0.295***		-0.299
					(0.0626)		(0.247)
Ln_Electricity Price						-0.245***	-0.113*
						(0.0447)	(0.0668)
Time Trend	0.0426***	0.0463***	0.0626***	0.0447***	0.0371***	0.0414***	0.0407***
	(0.00394)	(0.00426)	(0.00736)	(0.00427)	(0.00420)	(0.00397)	(0.00926)
Time Trend Squared	-0.000553***	-0.000585***	-0.000902***	-0.000614***	-0.000466***	-0.000591***	-0.000405**
	(7.14e-05)	(7.41e-05)	(0.000117)	(7.61e-05)	(7.95e-05)	(7.24e-05)	(0.00016)
Constant	6.421***	6.365***	6.325***	5.945***	6.538***	6.413***	8.540***
	(0.150)	(0.154)	(0.190)	(0.0841)	(0.203)	(0.153)	-0.766
Observations	52	52	52	52	52	52	52
R-squared	0.799	0.783	0.744	0.774	0.770	0.793	0.847

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Authors calculations.

Based on the results, the own-price elasticity of jet fuel is negative, complying with a theory: as the price of jet fuel (big airplanes) goes up by 1%, it's quantity consumed is decreasing by 0.16%. All other fuels used in transportation sector, like distillate fuel (diesel in cars/buses/ trains/ship), HGL (propane in cars/buses), aviation fuel (small airplanes), motor gasoline (cars, motorcycles, light trucks, and boats), electricity (electric vehicles), demonstrate strong complementary relationship with jet fuel when regressed separately (Models 1-6): as their prices increase by 1% the jet fuel volume consumed decreases with a varying magnitude in the range of 0.1-0.3%. So, aviation gasoline, distillate fuel and motor gasoline seem to be strong complements to jet fuel (probably because of main focus of refineries), along with having a strong complementarity with HGL (propane) and electricity (probably produced more because of higher demand on those fuels).

The analysis reveals that the demand for jet fuel is highly inelastic to both its own price and cross-price variations in fuels such as distillate fuel, HGL (propane), aviation gasoline, motor gasoline and electricity. In other words, fluctuations in these fuel prices have a limited effect on the quantity of jet fuel consumed. This finding also underscores jet fuel's role as a complementary fuel in the overall transportation market, primarily supporting large-scale aviation operations. When the prices of alternative transportation fuels rise, the higher costs are likely passed on to consumers as more expensive airline tickets, leading to reduced air travel demand. Furthermore, the time trend in all models shows a steady increase in jet fuel usage over time, although the rate of growth is gradually slowing. This suggests that while jet fuel consumption continues to rise, its growth trajectory is becoming less pronounced.

3.5.1.4. Motor Gasoline

The results of regression with *Motor Gasoline Fuel* as a dependent variable are presented in Table-3.6.

Table-3.6. Substitutability between Dependent variable: *Ln_Motor Gasoline Volume* and other fuels in General Transportation sector in California in 1970-2021

VARIABLES	(1) Model-1	(2) Model-2	(3) Model-3	(4) Mode- 4	(5) Model-5	(6) Model-6	(7) Model-7
Ln _Aviation Gasoline Price	-0.0367						-0.359***
	(0.0276)						(0.113)
Ln_Distillate Fuel Price		-0.0402					-0.466***
		(0.0310)					(0.101)
Ln_HGL Price			-0.0155				0.0119
			(0.0418)				(0.0576)
Ln_Jet Fuel Price				-0.0104			0.410***
				(0.0180)			(0.109)
Ln_Motor Gasoline Price					-0.0111		0.157
					(0.0328)		(0.11)
Ln_Electricity Price						-0.0526**	-0.0467
						(0.0235)	(0.0298)
Time Trend	0.0272***	0.0277***	0.0279***	0.0271***	0.0267***	0.0272***	0.0293***
	(0.00217)	(0.00226)	(0.00366)	(0.00225)	(0.00220)	(0.00209)	(0.00414)
Time Trend Squared	- 0.000351** *	- 0.000355** *	- 0.000369** *	- 0.000356** *	- 0.000350** *	- 0.000358** *	- 0.000293** *
	(3.93e-05)	(3.93e-05)	(5.81e-05)	(4.01e-05)	(4.16e-05)	(3.81e-05)	(7.01E-05)
Constant	7.078***	7.074***	7.007***	6.994***	7.008***	7.146***	7.969***
	(0.0824)	(0.0818)	(0.0943)	(0.0443)	(0.106)	(0.0807)	(0.342)
Observations	52	52	52	52	52	52	52
R-squared	0.868	0.868	0.863	0.864	0.863	0.876	0.934

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Authors calculations.

According to the results, electricity demonstrates a complementary relationship with motor gasoline (cars, motorcycles, light trucks, and boats) when regressed separately (Models 1-6): as its price increases by 1% the motor gasoline volume consumed decreases by 0.05% accordingly. So, electricity tends to have a strong complementarity with motor gasoline, probably because of higher demand for electric vehicles in past years.

The results also show that the demand for motor gasoline seems to be highly inelastic to cross-price changes in electricity, meaning that fluctuations in the prices of this fuel have a minimal impact on the quantity of motor gasoline consumed. The results also highlight motor gasoline's complementary role in the fuel market, particularly with electric vehicles, as it remains a primary energy source for cars, motorcycles, light trucks, and boats. When the costs of electric vehicle fuels rise, there is a tendency for reduced usage of motor gasoline-driven vehicles, likely reflecting a shift in transportation choices. Additionally, similarly as in previous models, the time trend across models shows that motor gasoline consumption has been increasing over time, but the growth rate is gradually slowing, indicating a deceleration in its usage expansion.

3.5.1.5. Electricity

The results of regression with *Electricity* as a dependent variable are presented in Table-3.7.

Table-3.7. Substitutability between Dependent variable: *Ln_Electricity Volume* and other fuels in in General Transportation sector in California in 1970-2021

VARIABLES	(1) Model-1	(2) Model-2	(3) Model-3	(4) Mode- 4	(5) Model-5	(6) Model-6	(7) Model-7
Ln _Aviation Gasoline Price	0.220*						-1.548***
	(0.113)						(0.478)
Ln_Distillate Fuel Price		0.342***					-0.692
		(0.123)					(0.43)
Ln_HGL Price			0.392**				0.077
			(0.166)				(0.244)
Ln_Jet Fuel Price				0.203***			0.75
				(0.0699)			(0.461)
Ln_Motor Gasoline Price					0.456***		1.430***
					(0.121)		(0.468)
Ln_Electricity Price						0.0840	0.042
						(0.103)	(0.127)
Time Trend	0.0883***	0.0830***	0.0632***	0.0844***	0.0945***	0.0901***	0.106***
	(0.00892)	(0.00892)	(0.00892)	(0.00892)	(0.00892)	(0.00892)	(0.0176)
Time Trend Squared	-0.000947***	-0.000917***	-0.000538**	-0.000880***	-0.00109***	-0.000927***	-0.00110***
	(0.000162)	(0.000162)	(0.000162)	(0.000162)	(0.000162)	(0.000162)	(0.0003)
Constant	-1.835***	-2.069***	-2.062***	-1.620***	-2.645***	-1.484***	-1.395
	(0.339)	(0.324)	(0.375)	(0.172)	(0.392)	(0.353)	(1.454)
Observations	52	52	52	52	52	52	52
R-squared	0.893	0.900	0.896	0.902	0.911	0.886	0.943

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Authors calculations.

Based on the results, aviation gasoline (small airplanes), jet fuel (big airplanes), motor gasoline (cars, motorcycles, light trucks, and boats), distillate fuel (diesel in cars/buses/ trains/ship) and HGL (propane in cars/buses) demonstrate a substitution relationship with electricity (electric vehicles) when regressed separately (Models 1-6): as their prices increases by 1% the electricity volume consumed increases with a varying magnitude range of 0.1-0.4%, accordingly.

The findings reveal that electricity demand is largely unaffected by cross-price variations in aviation fuels, motor gasoline, distillate fuels, and HGL, reflecting its inelastic nature. This suggests electricity has traditionally served as an alternative in the fuel market, especially for transportation modes like aviation, conventional vehicles, and ships, with its primary use being in electric vehicles. Consequently, as the costs associated with aviation/jet fuels, motor gasoline, diesel/propane or ship travel increase, there is a shift toward greater adoption of electric vehicles. Furthermore, the time trend reveals a steady increase in electricity consumption over time, though the rate of growth has been gradually slowing.

3.5.2. Summary of Results on Air Transportation Fuels Substitutability with Other Fuels in General Transportation Sector in California

This part of paper summarizes the results from analyzing the historical relationships between fuel usage in aviation and other transportation sectors (cars, ships, trains, buses, trucks, etc.) in general transportation sector in California for 1970-2021. Overall, the results show that aviation (through aviation gasoline and jet fuel) plays a complementary role in the transportation fuels market: when it becomes more expensive to travel by other means of transportation, so prices of other fuels increase (cars/buses/trucks/ships/trains, etc.), people prefer to fly less on small/big jets, probably because of higher/more expensive costs for airplane tickets in general.

Aviation fuels, however, appear to relate to other fuels differently: both aviation fuels seem to have a substitute relationship with HGL (propane) and electricity use: higher are costs for flying on airplanes in general, the more people would prefer to travel by cars, buses, ships and electric vehicles.

One of the reasons for this substitution relationship with different types of fuels is because of production complementarity in refineries: most refineries in the U.S. focus on producing transportation fuels like motor gasoline, distillate fuel and jet fuel/aviation gasoline fuel in highest volumes simultaneously, while producing *in addition* other fuels along with market demand (HGL, electricity). Also, aviation historically has been a substitution for cars/buses (propane and electric) and ship travel. That's why both aviation fuels (aviation gasoline and jet fuel) have strong substitution with HGL (propane) and electricity. That's because of the primary focus of refineries on producing aviation/jet fuels, distillate (diesel) fuel, motor fuel and because of historical alternative option for travelling (Ref: https://www.eia.gov/energyexplained/oil-and-petroleum-products/refining-crude-oil.php).

The analysis reveals that the demand for aviation gasoline and jet fuel is very inelastic with respect to the own-price and cross-price changes in distillate fuel, HGL (propane), motor gasoline and electricity: quantity of aviation gasoline/jet fuel changes slower than price changes in above-mentioned fuels.

Even those results are well-known, we did this analysis using historical EIA data to show the relationship between aviation and other transportation areas in general when biofuels (ethanol, biodiesel, renewable diesel, etc.) are already introduced to transportation market (in 2010's), and to see how it changes (if any) inter-substitution/complementarity inside transportation sector with respect to aviation, in general, before looking into more detailed analysis of biofuels use particularly in aviation sector. We detect that introducing the biofuels from 2010's does not seem to change much the traditional substitution and complementarity relationships between aviation fuels and other transportation fuels: when it becomes more expensive to travel by other means of transportation (cars/buses/trucks/trains, etc.), people prefer to fly less, probably because of more expensive costs for airplane tickets in general, but when there are higher costs for flying on airplanes in general, people would prefer to travel more by cars, buses and electric vehicles.

3.5.3. Results on Aviation Fuels Substitutability in Air Transportation sector in California

As discussed above, in this second part of the analysis, the competition results between jet fuel and biomass-based SAF in air transportation sector in California are presented. The correlation matrix demonstrates no major correlation between variables (Appendix-C: Table-C2).

3.5.3.1. Basic Static Model Results

The results of Basic Static Model are presented in Table-3.9. The data is used for the period of 2013-2022 (full dataset).

Table-3.9. Results: Basic Static Model

VARIABLES	(1) Model 1- LN Jet Fuel Volume	(2) Model 2- LN SAF Volume
LN Jet Fuel price	0.199***	1.376***
	(0.0622)	(0.227)
LN SAF Price	-0.0492	0.534**
	(0.0579)	(0.211)
Constant	16.03***	11.38***
	(0.0578)	(0.211)
Observations	111	111
R-squared	0.097	0.273

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Authors calculations.

The results from Model-1 reveal that when jet fuel is defined as a dependent variable, the jet fuel's own price elasticity is significantly positive: 1% increase in jet fuel price leads to 0.19% increase in jet fuel volumes sold. This result contradicts a bit with a theory, where the own-price elasticity is expected to be negative. The cross-price elasticity of SAF price on jet fuel volume seems not to be significant, which means that any increase in SAF price does not affect the jet fuel volumes sold. This result may align with the fact that SAF market is substantially smaller than jet fuel market and SAF market constitutes only 2% of whole fuels market, according to the data analyzed (EIA, EPA).

On the other hand, the results from Model-2 reveal that when SAF is defined as a dependent variable, the SAF's own price elasticity is significantly positive with a higher magnitude: 1% increase in jet fuel price leads to 0.53% increase in SAF volumes sold. This result contradicts again with a theory, where the own-price elasticity is expected to be negative. Also, the cross-price elasticity of jet fuel price on SAF volume seems to be significant: 1% increase in jet fuel price leads to 1.37 % increase of SAF volumes sold. This result demonstrates that SAF is a substitute for jet fuel volume consumed.

3.5.3.2. Alternative Model using Time Trend Results

The results of the alternative model using time trend are presented in Table-3.10. The data is used for the period of 2013-2019 (restricted dataset). The reason for using the restricted dataset of 2013-2019 data is to capture inference before COVID period, without the *unusual* effect of pandemic on fuels market.

Table-3.10. Results: Alternative Model with Time Trend

VARIABLES	(1) Model 1-LN Jet Fuel Volume	(2) Model 2-LN SAF Volume
LN Jet Fuel price	0.0465	1.043***
	(0.0426)	(0.338)
LN SAF Price	0.0146	0.216
	(0.0298)	(0.237)
Time Trend	0.00600***	-0.0416***
	(0.00196)	(0.0155)
Time Trend Squared	-3.32e-05	0.000288*
	(2.07e-05)	(0.000164)

Constant	16.07***	12.61***
	(0.0725)	(0.575)
Observations	84	84
R-squared	0.567	0.674

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Source: Authors calculations.

The results from Model-1 reveal that when jet fuel is defined as a dependent variable, the jet fuel's own price elasticity is significantly positive, but with lower magnitude than in an initial basic static model: 1% increase in jet fuel price leads to 0.04% increase in jet fuel volumes sold. This result still contradicts a bit with a theory, where the own-price elasticity is expected to be negative. The cross-price elasticity of SAF price on jet fuel volume seems not to be significant again in this model, which means that any increase in SAF price does not correlate to the jet fuel volumes sold. The reason is probably because SAF market is substantially smaller than jet fuel market and SAF market constitutes only 2% of whole air fuels market, according to the data analyzed (EIA, EPA). The coefficient on the time trend variable is significantly positive, which means that the percentage growth rate of jet fuel volumes sold per period is positive (upward trending slope).

On the other hand, the results from Model-2 reveal that when SAF is defined as a dependent variable, the SAF's own price elasticity is not any more significant, even though still positive. Also, the cross-price elasticity of jet fuel price on SAF volume is estimated to be highly statistically significant as before, but with lower magnitude: 1% increase in jet fuel price leads to 1.043% increase of SAF volumes sold. Even though the magnitude of this impact is small (around 1.0%), it is still positively significant as in the initial basic model (1.37%) and demonstrates that SAF is a substitute for a jet fuel volume consumed. The coefficient on the time trend variable is significantly negative, while that on the time trend squared is positive, all with low magnitudes, which means that the percentage growth rate of SAF volumes consumed per period is slightly decreasing but with an increasing rate (downward trending slope). In addition, the results show that I)the demand for SAF is elastic with respect to the cross-price changes in kerosine-based jet fuel, indicating that quantity of SAF changes more than price changes of jet fuel; ii) and it is inelastic with respect to the own-price changes in SAF: quantity of SAF changes slower than its price changes.

3.5.3.3. Alternative Model using Seemingly Unrelated Regression Results

Given that both models use the same independent variables and to check if the specification with including time trend and using OLS regression is justified, SUR is used, hypothesizing that there might be a correlation in error terms between two models: jet fuel volume model and SAF volume model (used as dependent variables). The results of the alternative regression using SUR are presented in Tables-3.11- 3.12. The data is used for the period of 2013-2019 (restricted dataset) as in the above model using time trends.

Table-3.11. Results: Alternative Model using Seemingly Unrelated Regression

VARIABLES	(1) Model 1-LN Jet Fuel Volume	(2) Model 2-LN SAF Volume
LN Jet Fuel price	0.0465	1.043***
	(0.0422)	(0.396)
LN SAF Price	0.0146	0.216
	(0.0302)	(0.203)
Time Trend	0.00600***	-0.0416**
	(0.00222)	(0.0174)
Time Trend Squared	-3.32e-05	0.000288
	(2.33e-05)	(0.000175)
Constant	16.07***	12.61***
	(0.0766)	(0.689)
Observations	84	84
R-squared	0.567	0.674

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Authors calculations.

Table-3.12. Correlation Matrix of Residuals and Breusch–Pagan test of independence

VARIABLES	Model 1-LN Jet Fuel Volume	Model 2-LN SAF Volume
Model-1: LN Jet Fuel Volume	1.0000	
Model-2: LN SAF Volume	0.0667	1.0000

Source: Authors calculations.

Breusch–Pagan test of independence	H_0: No correlation between two models
Chi2(1) =	0.374
P-value	0.5408

Source: Authors calculations.

The results from SUR demonstrate that both models used in regression have the same coefficient estimates and still have substitution effect of jet fuel prices on SAF volume sold but demonstrates higher standard errors than above alternative OLS specification with time trend. Despite revealing very small correlation in the residuals between the jet fuel volume and SAF volume equations, the Breusch–Pagan test of independence finds that *"no correlation"* between the two equations cannot be rejected. The OLS specification in previous part is then determined to be more efficient.

3.5.4. Summary of Results on Aviation Fuels Substitutability in Air Transportation sector in California

This part of paper summarizes the results from analyzing the substitutability between traditional jet fuel and novel SAF in air transportation sector in California for 2013-2019. Overall, the results from this part of the paper show that when traditional kerosine-based jet fuel volume is defined as a dependent variable, the cross-price elasticity of SAF price on jet fuel volume seems not to be affecting, which means that any increase in SAF price does not affect the jet fuel volumes sold. The reason for this might be the SAF market size: it is substantially smaller than the jet fuel market and SAF market constitutes only 2% of whole air fuels market (EIA, EPA). The time trend shows that the growth rate of jet fuel volumes consumed per period is increasing over time. So, biofuels in aviation are too small yet to affect the traditional kerosine-based jet fuels market, which proves the other recent literature findings in the field.

On the other hand, when SAF is defined as a dependent variable, the cross-price elasticity of jet fuel price on SAF volume seems to be highly significant: 1% increase in jet fuel price leads to 1.043% increase of SAF volumes consumed. Even the magnitude of this impact is small (around 1.0%), but it is still positively significant and demonstrates that SAF is a valid substitute for jet fuel volume consumed. Furthermore, the time trend variable is significantly negative at an increasing rate. We should consider the analyzed data period of 2013-2019: at this period the SAFs have been just appearing in the market and they were in their initial development stage yet because of their novelty. Based on the highly increased dynamics of SAF volumes consumed during 2020-2022 period (Fig.-3.4), it is highly possible that SAFs consumption would show increasing trend in further analysis for more recent years (after 2022), when data becomes available.

3.6. Conclusion and Policy Implications

The transportation sector in California presents significant potential for substituting biomass for fossil fuels, being the largest biomass consumer in the U.S. Biomass is recognized as an energy source that helps reduce CO_2 emissions. However, the sector faces challenges from the volatility of fossil fuel prices. The ability of the transportation sector to adapt to these potential challenges will depend on its capacity for inter-fuel substitution given its technological framework. Particularly, this relates to the aviation sector, where the introduction of biofuels from 2010's leads to impressive development and application of novel energy sources that reduces emissions in aviation. The economic costs of energy and environmental policies in aviation also depend on the substitutability of biomass and traditional kerosine-type fuels in aviation industry.

Accordingly, this study contributes to the literature 1) by examining the substitution or complementarity of different fuels consumption in transportation sector in general, after the introduction of biofuels (2013-2021), but not separating biofuels into a separate category because of data limitation and 2) by examining the extent to which the use of biofuels contributes to inter-fuel substitution particularly in aviation sector through price elasticities. The estimation results of the series of OLS models including time trends provide additional insights on inter-fuel substitution and biomass use in the transportation sector in general and in aviation industry in particular.

The results from the first part of the paper, where general transportation sector is analyzed, shows that aviation (through aviation gasoline and jet fuel) plays a complementary role in the transportation fuels market: when it becomes more expensive to travel by other means of transportation (cars/buses/trucks/trains, etc.), people prefer to fly less on small/big jets, probably because of higher/more expensive costs for airplane tickets and also probably because of production complementarity in refineries, who prefer to focus on producing mostly aviation fuels (aviation gasoline and jet fuel), motor gasoline and distillate fuel in larger amounts *simultaneously* because of higher demands/profits from those types of fuels (producing other fuels as demanded). Another important finding is that both aviation fuels (aviation gasoline and jet fuel) seem to have a substitute relationship with HGL (propane) and electricity (use of aviation is a substitute to use of cars/buses, ships, electric vehicles) even after introduction of the biofuels into transportation sector, complying with previous studies. The reason for that is probably because of historical substitution relationship between aviation and travel by cars (propane, electric) and ships. In addition, the results show that demands for aviation gasoline and jet fuel are very inelastic with respect to the own-price and cross-price changes in distillate fuel, HGL (propane), aviation gasoline/jet fuel and electricity. So, introduction of biofuels seems not to change traditional relationships between different transportation fuels in general.

The results from the second part of the paper, where the aviation industry is analyzed, show that 1) any changes in SAF price does not affect the jet fuel volumes consumed, due to the fact that biofuels in aviation are too small yet to affect the traditional kerosine-based jet fuels market; 2) changes in traditional kerosine-based jet fuel price affects positively the SAF volumes consumed, meaning that SAF becomes a valid substitute for a traditional kerosine-type jet fuel use in aviation. In addition, the results show that the demand for SAF is elastic with respect to the cross-price changes in kerosine-based jet fuel, and inelastic with respect to the own-price changes in SAF.

In general, introduction of biofuels into transportation sector does not change traditional substitution and complementary relationships between different fuel types inside general transportation sector yet, but already becomes a valid substitute to traditional kerosine-type jet fuel use in the aviation sector in California, in particular.

The findings are of importance to policy makers because they provide critical information about the results of governmental programs that incentivize the use of biofuels to reduce CO_2 emissions in California, particularly in the aviation industry[7]. As a further step, the analysis of biofuels impact on CO_2 emissions levels in California would be beneficial for more detailed analysis of inter-fuel substitution in the region.

[7] While this chapter focuses primarily on liquid biofuels such as Sustainable Aviation Fuel (SAF), it is worth noting that emerging research also explores solar-based synthetic fuels—produced using concentrated solar power to generate syngas—as a future option for decarbonizing aviation (see https://caseyhandmer.wordpress.com/2022/02/03/terraform-industries-whitepaper/).

REFERENCE

1. Anderson, S. T. (2012). The demand for ethanol as a gasoline substitute. *Journal of Environmental Economics and Management, 63*(2), 151-168. https://doi.org/10.1016/j.jeem.2011.08.002
2. Bardazzi, R., & Pazienza, M. G. (2024). Decarbonising transport: Can we rely on fuel taxes? *Transportation Research Part D: Transport and Environment, 136*, 104391. https://doi.org/10.1016/j.trd.2024.104391
3. Bartocci, A., & Pisani, M. (2013). "Green" fuel tax on private transportation services and subsidies to electric energy. A model-based assessment for the main European countries. *Energy Economics, 40*(S1), 32-57. https://doi.org/10.1016/j.eneco.2013.09.019
4. Basso, L. J., Lin, C., & Snyder, L. V. (2007). A study on the fuel demand model for the U.S. transportation sector. *Energy Economics, 29*(4), 712-725.
5. Brito, T. L. F., Islam, T., Mouette, D., Meade, N., & dos Santos, E. M. (2020). Fuel price elasticities of market shares of alternative fuel vehicles in Brazil. *Transportation Research Part D: Transport and Environment, 89*, 102643. https://doi.org/10.1016/j.trd.2020.102643
6. 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
7. Center for Law, Energy & the Environment. (2022, October). *Clean Takeoff: Policy solutions to promote sustainable aviation in California*. University of California, Berkeley. https://www.law.berkeley.edu/wp-content/uploads/2022/10/Clean-Take-Off-2022.pdf
8. Chen, X., Huang, H., Khanna, M., & Önal, H. (2014). Alternative transportation fuel standards: Welfare effects and climate benefits. *Journal of Environmental Economics and Management, 67*(3), 241-257. https://doi.org/10.1016/j.jeem.2013.09.006
9. Cherubini, F. (2011). *Applied econometrics with Stata: A practical guide to econometric analysis*. Palgrave Macmillan.
10. Edelenbosch, O. Y., van Vuuren, D. P., Bertram, C., Carrara, S., Emmerling, J., Daly, H., Kitous, A., McCollum, D. L., & Saadi Failali, N. (2017). Transport fuel demand responses to fuel price and income projections: Comparison of integrated assessment models. *Transportation Research Part D: Transport and Environment, 55*, 310–321. https://doi.org/10.1016/j.trd.2017.03.005
11. Espey, M. (1998). Gasoline demand revisited: An international meta-analysis of elasticities. *Energy Economics, 20*(3), 273-291. https://ideas.repec.org/a/eee/eneeco/v20y1998i3p273-295.html
12. Hochman, G., & Timilsina, G. R. (2017). Fuel efficiency versus fuel substitution in the transport sector: An econometric analysis (Report No. 8070). The World Bank. https://documents.worldbank.org/pt/publication/documents-reports/documentdetail/451701495467642019/fuel-efficiency-versus-fuel-substitution-in-the-transport-sector-an-econometric-analysis
13. Hossain, A. K. M. N., & Serletis, A. (2020). Biofuel substitution in the U.S. transportation sector. *The Journal of Economic Asymmetries, 22*, 1-10. http://dx.doi.org/10.2139/ssrn.3577371
14. Huse, C. (2018). Fuel choice and fuel demand elasticities in markets with flexfuel vehicles. *Nature Energy, 3*, 582-588. https://ssrn.com/abstract=4314791

15. Kim, Y.-D., Han, H.-O., & Moon, Y.-S. (2011). The empirical effects of a gasoline tax on CO_2 emissions reductions from the transportation sector in Korea. *Energy Policy, 39*(2), 981-989. https://doi.org/10.1016/j.enpol.2010.11.026
16. Lin, C., Basso, L. J., & Snyder, L. V. (2013). Fuel demand modeling in the U.S. transportation sector: Revisiting price elasticities. *Energy Economics, 36*(1), 308-317.
17. McCollum, C. J., Ramsey, S. M., Bergtold, J. S., & Andrango, G. (2021). Estimating the supply of oilseed acreage for sustainable aviation fuel production: Taking account of farmers' willingness to adopt. *Energy, Sustainability and Society, 11*(1), 1-14. https://doi.org/10.1186/s13705-021-00308-2
18. Neves, S. A., & Marques, A. C. (2021). The substitution of fossil fuels in the US transportation energy mix: Are emissions decoupling from economic growth? *Research in Transportation Economics, 90*, 101036. https://doi.org/10.1016/j.retrec.2021.101036
19. Nugroho, Y. K., & Zhu, L. (2024). Strategic supply and transportation planning of a supply chain for agricultural biomass to hydrogen and syngas. *Computers & Industrial Engineering, 197*, 110640. https://doi.org/10.1016/j.cie.2024.110640
20. Shriver, S. K. (2015). Network effects in alternative fuel adoption: Empirical analysis of the market for ethanol. *Marketing Science, 34*(1), 78-97. https://doi.org/10.1287/mksc.2014.0881
21. Suh, D. H. (2019). Interfuel substitution effects of biofuel use on carbon dioxide emissions: Evidence from the transportation sector. *Applied Economics, 51*(31), 3413-3422. https://doi.org/10.1080/00036846.2019.1581906
22. Uchôa, F., de Jesus, C. S., & Cardoso, L. C. B. (2020). Fuel demand elasticities in Brazil: A panel data analysis with instrumental variables. *International Journal of Energy Economics and Policy, 10*(2), 450-457. https://www.econjournals.com/index.php/ijeep/article/view/8787
23. U.S. Energy Information Administration. (2024, June 28). *State Energy Data System (SEDS): 1960-2022, California Profile*. U.S. Energy Information Administration. Retrieved November 23, 2024, from https://www.eia.gov/state/seds/
24. U.S. Energy Information Administration. (2023). *Transportation sector energy consumption estimates, selected years, 1960-2022, California* (SEDS Report No. 1022). U.S. Department of Energy. Retrieved November 23, 2024, from https://www.eia.gov/state/seds/sep_use/tra/pdf/use_tra_CA.pdf
25. U.S. Energy Information Administration. (2023). *Transportation sector energy price and expenditure estimates, selected years, 1970-2022, California* (SEDS Report No. 1024). U.S. Department of Energy. Retrieved November 23, 2024, from https://www.eia.gov/state/seds/sep_prices/tra/pdf/pr_tra_CA.pdf
26. U.S. Environmental Protection Agency. (n.d.). Public data on the Renewable Fuel Standard. U.S. Environmental Protection Agency. Retrieved November 23, 2024, from https://www.epa.gov/fuels-registration-reporting-and-compliance-help/public-data-renewable-fuel-standard
27. Winchester, N., McConnachie, D., Wollersheim, C., & Waitz, I. (2013, March 1). *Market cost of renewable jet fuel adoption in the United States* [Report]. Partnership for Air Transportation Noise and Emissions Reduction. United States Department of Transportation, Federal Aviation Administration, Office of Environment and Energy. https://rosap.ntl.bts.gov/view/dot/28422
28. Wu, J. J., & Langpap, C. (2014). The price and welfare effects of biofuel mandates and subsidies. *Environmental and Resource Economics, 62*(1), 35-57. https://doi.org/10.1007/s10640-014-9814-8

29. Zhao, X., Taheripour, F., Malina, R., Staples, M. D., & Tyner, W. E. (2021). Estimating induced land use change emissions for sustainable aviation biofuel pathways. *Science of the Total Environment, 776*, 146238. https://doi.org/10.1016/j.scitotenv.2021.146238
30. Zellner, A. (1963). "Estimators for Seemingly Unrelated Regression Equations: Some Exact Finite Sample Results." *Journal of the American Statistical Association*, 58(304), 977-992. https://doi.org/10.1080/01621459.1963.10480681

Chapter-4:
“Motives of Migration and Remittances in Tajikistan”

4.1. Introduction

Tajikistan is a landlocked Central Asian country that covers 144,000 square kilometers. It is bordered by Kyrgyzstan on the north, China on the east, Afghanistan on the south, and Uzbekistan on the west and northwest (Fig.-4.1). Almost 93% of its territory is mountainous and only about 10% is suitable for cultivation. Its mountainous nature complicates transportation and communication between regions, but also provides rich natural endowments of minerals, hydropower potential, and water resources for irrigation. According to preliminary data from the State Statistical Agency (SSA), the total population of Tajikistan in 2013 was 8.06 million. Within the population structure, 26.4% live in urban areas and 73.6% live in rural areas. The gender distribution is almost equal, i.e., 50.1% male and 49.9% female, which is unchanged during 1999–2013 (Borzhikova, 2014).

The migration phenomenon in Tajikistan is deeply rooted in the country’s economic and historical context. Following the collapse of the Soviet Union in 1991 and a subsequent civil war (1992-1997), Tajikistan experienced significant economic hardship, driving large-scale labor migration, particularly to Russia, which remains the primary destination for Tajik migrants (Roberts & Jones, 2020; Abdulloev et al., 2014). This trend has continued into the present day, largely due to persistent issues such as high unemployment, limited economic opportunities, and low wages at home (World Bank, 2023). Approximately 10-15% of Tajikistan’s population, primarily young men, currently work abroad, with remittances constituting a significant portion of the country's GDP—around 30%, one of the highest shares in the world (Ratha et al., 2022).

Tajikistan's economy is not creating enough jobs for its rapidly growing labor force (ADB, 2019a). Fifty five percent of the working-age population is not economically active as they do not participate in the labor force (JICA, 2019). The GDP of Tajikistan was TJS 40.5 billion in 2013, at purchasing power parity, the per capita income was close to $2,200, which classifies it as a factor-driven economy under the Global Competitiveness Index (Tajikistan Statistics Agency, Macroeconomic Indicators, 2013).

After 30 years of independence, enduring poverty and instability, limited agricultural and industrial production, low wages, and a deeply corrupt economic system have contributed to Tajikistan's rise as a major exporter of labor as a substitute for domestic production of goods or services. Driven by a lack of economic opportunities in Tajikistan, more than 2 million citizens from Tajikistan travel to Russia for work each year, according to the Russian Ministry of Internal Affairs (2018). But the actual number of Tajiks in Russia may be much higher, with as many as 40 percent working illegally and therefore not appearing within the official statistics. Surveys indicate that 30 to 40 percent of households in Tajikistan have at least one member working abroad. Tajikistan is one of the most remittance-dependent countries in the world: the remittance inflow in Tajikistan reached its peak in 2013, amounting to $US3.7 billion and making up 42 percent of GDP (World Bank, 2017). Since 2005, Tajikistan has been ranked as one of the top countries in terms of remittances' share of GDP, which has been as high as 50 percent.

The patterns of migration in Tajikistan are complex. According to the Integral Human Development report (2024), Russia remains the top choice for Tajik emigrants, attracting more than 90% of them. This strong migration pattern stems from deep-rooted historical ties between the two countries, supported by frequent flights, the absence of visa requirements, and the presence of a well-established Tajik community. Many Tajiks also speak Russian, which eases their integration. Migration is largely seasonal, with workers heading to Russia in the spring and coming back to Tajikistan by summer's end. For more than 25 years, Tajiks have mainly migrated to Russia. In 2023, 80% of remittances from Tajik migrants were sent from Russia. Migration from Tajikistan is male dominated. The largest group of migrants is 15–29 years old (45.4%) followed by the group aged 30–44 (39.5%) (JICA, 2019). IOM (2003) identified that 75% of Tajik migrants spoke Russian in addition to their native language; however, knowledge of a second or third language is more common among older people than the younger generation. The younger the migrant worker, the lower the level of their language skills. More than 85% of the labor migrants were from rural areas (JICA 2018). Roughly 75% of the migration is seasonal, meaning that the migrants return to Tajikistan at least once a year (JICA, 2018). Tajik migration of seasonal workers often occurs in groups. Migration networks are an important factor in labor migration in Tajikistan; they impact migration magnitude, job search, adaptation, and sending remittances (JICA, 2019). For 58% of the surveyed returned migrants, the skills (work experience and training) acquired in their home country were useful in finding employment. A majority of the migrant men find work in the construction sector primarily as unskilled laborers. The majority of women who migrate find work in the service sector. The migrant workers chose to work mainly in construction (59%), followed by trade and services (17%), manufacturing (5%), transport and communication (5%), and other (14%) sectors. These tend to be lower-paying positions. The types of job migrants engage in depend largely on the migrants' social capital (JICA 2019). The remittances to Tajikistan help the migrants' families pay for their daily expenses and immediate needs. Migrant families use an overwhelming portion of remittances (94%) for private consumption, which is the most important component of GDP. However, remittances are rarely used for investments or to save for future contingencies (ILO, 2010).

Figure-4.1. Political Map of Tajikistan

Source: https://www.nationsonline.org/oneworld/map/tajikistan-political-map.htm, 2019.

Labor migration has become both a solution and a challenge for Tajikistan. Remittances play a vital role in poverty reduction, allowing households to improve their quality of life, invest in healthcare, education, and housing, and reduce income disparities. However, high reliance on remittances has created vulnerabilities, leaving Tajikistan exposed to economic fluctuations in Russia and changes in Russian migration policies (Abdulloev et al., 2014; Brown et al., 2021). This dependence also has social repercussions, such as family separation, altered family roles with increased responsibilities for women, and a "brain drain" where skilled workers leave the country, limiting local development potential (Yormirzoev, 2017; Lee, 2023).

Today, Tajikistan faces the challenge of managing the impacts of migration. Economic diversification and local job creation are widely advocated as necessary steps to reduce dependency on foreign labor markets and support sustainable development (Roberts & Jones, 2020). Investments in education, vocational training, and policy frameworks that encourage entrepreneurship could gradually alleviate the pressures of labor migration and promote economic resilience within Tajikistan (World Bank, 2023).

Despite of being analyzed in many previous studies, the phenomenon of Tajikistan migration still experiences the lack of literature that empirically analyzes the motives of migration from the angles of seminal well-known theoretical models that leads to migration. For example, there are many studies, as cited in the next section, that highlights different patterns of migration in Tajikistan: driving factors of migration from Tajikistan, the importance of networks and communities of migrants for migration decisions, economic hardship as "push" factor for migration, low-skills jobs sector that is mainly occupied with migrants, remittances that cover mainly the substance costs of families of migrants in Tajikistan. But there is still a shortage of literature that empirically tests the well-known theoretical economic models that predict those patterns.

Also, majority of these studies do not consider the issue of endogeneity when dealing with decisions about migration and remittances: migrants and remitters are not drawn from a randomly selected sample population, but from individuals who self-select into these activities. Thus, migration is likely to be correlated with factors that affect remittances decisions of the households.

The purpose of this paper is to empirically analyze what factors drive people to migrate from Tajikistan and why do they remit, based on predictions of the seminal theoretical economic models. Also, this paper aims to correct the self-selection bias that might be present in the decisions of migrating or remitting on household level data in Tajikistan.

Based on the Probit and Heckman correction methods, we find that migration cost variables (having networks, family arrangement of costs related to migration) have a strong positive association with the probability of having migrants in the households, matching the Borjas (1987, 1991) and CHW (2007) models that predict migration rate increasing due to migration costs decreasing. The association is stronger in rural or urban areas depending on analyzed migration cost variable. Also, results show that education and skills variables are negatively related to the probability of having migrants in the households, especially in urban areas, which supports the Negative Selection migration process predicted by Roy's model (1951). Language skills, family size and family income variables demonstrate a positive association with remittances sent. The results also reveal that there exists a significant selection bias due to non-randomness of migration or remittances decisions and therefore Heckman estimation method usage is justified to correct for the self-selection bias.

The rest of the paper is organized as follows. In the next sections literature review, theoretical frameworks, estimation strategy and data sections are outlined. Then results section presents the main results and discussions. The last section concludes.

4.2. Literature Review

While previous studies reveal general unconditional patterns regarding migration in Tajikistan (JICA 2019), less is known about the marginal impacts of determinants in a conditional sense. The literature is suggestive of nuances to the story of migration and remittances that has not yet been explored for the case of Tajikistan.

Migration and remittances are intricately connected phenomena that significantly influence global economies and individual livelihoods. As individuals move in search of better opportunities, they often maintain strong ties to their home countries, resulting in substantial financial transfers known as remittances. This review explores the theoretical frameworks that explain migration patterns, including key contributions, as well as the empirical research that illuminates the dynamics of migration and remittances.

4.2.1. Determinants of Migration

The seminal theoretical model by Roy (1951) introduced the concept of self-selection in migration, suggesting that individuals migrate based on the expected economic benefits they anticipate in the destination country. According to the Roy model, migration decisions are influenced by individuals' comparative advantages in different labor markets. Those with higher skills or productivity in specific sectors are more likely to move to places where their abilities are in demand, resulting in a sorting effect within the labor market. This model highlights that migrants are not homogenous; their potential economic contributions vary depending on their skills and the labor market conditions of the destination country.

Building on Roy's work, Borjas (1987, 1991) examined the influence of income inequality and labor market structures on migration selection. Borjas showed that nations with higher income inequality tend to attract more high-skilled migrants, as the returns on skill are greater. In contrast, countries with more equal income distributions may draw lower-skilled individuals. His research also underscores the significant role of immigration policies in shaping migrant flows and the skill composition of incoming workers, forming the foundation for extensive empirical studies on immigrant selection and labor market outcomes.

Rapoport and Docquier (2005) investigated the effects of migration on sending countries, with a particular focus on the "brain drain" phenomenon and the benefits of remittances. While acknowledging that the emigration of high-skilled workers can lead to a loss of talent in source countries, they also pointed out that remittances sent by migrants can enhance welfare and stimulate human capital investment in the countries of origin. Their work bridges the gap between the individual decisions of migrants and the broader economic impacts on both sending and receiving nations.

Clark, Hatton, and Williamson (2007) provided a historical perspective on migration, particularly the transatlantic migration from Europe to North America during the 19th and early 20th centuries. Their analysis emphasized factors such as wage differentials, population pressures, and the role of immigration policies, showing how changes in "push" and "pull" factors affected migration trends. This research reinforces the importance of economic opportunities, migration costs, and demographic conditions as key determinants of migration.

Economic difficulties, including poverty, unemployment, and inequality, are prominent push factors driving migration. Dao et al. (2018) found that while financial constraints are significant for the poorest countries, their impact is limited. Instead, a substantial portion of the growing segment can be attributed to changes in skill composition and macroeconomic factors, which tend to remain stable in the short term. Czaika and Vothknecht (2014) argue that migrants tend to have higher aspirations prior to migration, influenced by factors like youth, education, and socio-economic background. While migration generally benefits migrants economically, it also appears to increase their economic aspirations, trapping them on a "hedonic treadmill." Grogger and Hanson (2011) suggest that migrants are more educated than non-migrants in source-destination pairs where the earnings gap between the two countries is larger. Sorting patterns show that the share of more educated migrants in a destination increases with the earnings difference between high- and low-skilled workers. Naudé's (2010) research on migration trends from 45 Sub-Saharan African countries (1965–2005) highlights conflict and unemployment as key drivers of emigration. His findings suggest that prolonged conflict significantly increases outward migration, while weaker economic growth tends to reduce it. Although demographic and environmental challenges have limited direct influence, they may indirectly shape migration patterns by contributing to instability and lack of job opportunities. The study also identifies a migration hump, indicating that much of the region's migration is driven by forced rather than voluntary factors. Amare (2021) finds that urban growth in Nigeria, measured by night light intensity, encourages youth migration, especially among educated youth and women. Migration patterns vary by group and asset ownership—land encourages temporary moves, while livestock discourages migration. The study calls for targeted, multifaceted policies to manage rural–urban youth migration effectively.

Social networks play a crucial role in shaping migration decisions. Massey et al. (2013) showed that networks reduce the costs and risks of migration, making it easier for individuals to migrate, especially for first-time migrants. Liu (2013) demonstrated that having more resources as a result of strong ties appears to dampen overall migration, while having more resources as a result of weaker ties appears to stimulate male migration. Ruyssen et al. (2014) explore the under-researched area of South-South migration by examining intraregional migration in Sub-Saharan Africa through an extended human capital model that includes spatial dynamics. Analyzing bilateral data from 1980 to 2000, they find that migration within the region is mainly influenced by economic prospects and political conditions in destination countries, with proximity playing a key enabling role. The study also highlights the importance of migrant networks and environmental factors, emphasizing that spatial connections between origin and destination areas are crucial in understanding migration flows.

4.2.2. Determinants and Motives of Remittances in Developing Countries

Remittances are a critical financial resource for families in developing countries, contributing significantly to national economies. Recent studies have explored the economic, social, and personal factors that influence remittance behavior, as well as the various determinants that shape the volume and frequency of these remittances.

Economic factors are a major driving force behind remittances. Adams and Cuecuecha (2010) found that migrants often remit to meet the financial needs of their families, especially in countries with limited social security or welfare systems. Migrants typically send money home to cover essential expenses such as food, housing, education, and healthcare. Yang (2011) examined the impact of exchange rates and income levels in host countries, noting that favorable exchange rates and higher earnings among migrants tend to result in larger remittance flows, indicating how economic conditions in both the host and sending countries influence remittance patterns.

Many migrants are also motivated to remit out of altruism or a sense of family obligation. Carling (2008) found that family pressure and societal expectations often influence remittance behavior. In many cultures, migrants feel a moral duty to support their relatives, especially when they are dependent on this financial assistance for maintaining their livelihoods.

Remittances are increasingly being used to finance education and investments in human capital. Acharya, and Leon-Gonzalez (2014) found remittances enable severely credit-constrained households to enroll children in school and prevent dropouts. For households with less severe liquidity constraints, remittances boost investment in higher-quality education. Dinkelman and Mariotti (2016) found that circular labor migration impacts origin communities by boosting the human capital of the next generation: human capital higher among cohorts in communities with better access to migrant jobs.

Additionally, remittances are often used for investment in assets, businesses, or savings in the home country, as migrants view these transfers as a long-term financial strategy. Amuedo-Dorantes (2023) highlighted how remittances can support investments, particularly in real estate and small enterprises. Balli and Rana (2015) noted that remittances also serve as a form of risk diversification, allowing migrants to invest in their home country while reducing their exposure to economic risks in the host country. Wouterse (2010) identifies key characteristics of households that engage in international migration in rural Burkina Faso. These households tend to be larger, older, and possess more capital, such as land, livestock, and equipment. Remittances sent by international migrants are substantial, allowing receiving households to invest in livestock, which serves as both capital and a form of insurance. Wouterse also outlines three primary effects of migration on sending communities: remittances, which are allocated privately by migrant families; returns, with migrants bringing back new skills and ideas; and recruits, where households with migrants are more likely to send additional members, potentially reducing inequality and fostering economic improvement in the broader community.

Finally, remittances function as informal insurance, helping families cope with economic shocks such as illness, crop failures, or natural disasters. Yang and Choi (2007) found that remittance flows tend to rise in response to adverse economic conditions in the home country, indicating that migrants provide financial support as a safety net during periods of uncertainty.

Chami et al. (2009) found that remittances reduce output growth volatility in recipient countries, acting as a stabilizing force. The decline in remittances due to the global financial crisis could increase output variability, posing a challenge for governments that must rely on limited policy tools, like fiscal policy, to mitigate the negative economic and social effects.

4.2.3. Migration and Remittances in Tajikistan

Migration and remittances are crucial elements of Tajikistan's economic landscape, significantly impacting its national development and household welfare. Since the country gained independence in the early 1990s, emigration has been a consistent trend, driven by economic hardship and the search for better opportunities. Past empirical studies on Tajikistan are limited, with only few studies that analyzed the determinants of migration and remittances.

Using a household-level dataset from 2007 World Bank Living Standard Measurement Survey Tajikistan data, Abdulloev et al. (2011) empirically demonstrate that labor migration in Tajikistan serves as a substitute for informal economic activities. The authors argue that income from migration is an imperfect trade-off for income from informal work in Tajikistan. They highlight that individuals with professional skills are more likely to engage in informal activities locally, while those with lower skills or without secondary education tend to migrate, predominantly to Russia, in search of better job opportunities and higher earnings.

Meier’s (2014) study on low-skilled labor migration in Tajikistan explores the determinants driving this migration, particularly economic hardship and lack of job opportunities in the country by using World Bank 2007 Tajikistan Living Standard Survey (TLSS 2007) data. A Probit model marginal effects reveal that variables like working age of household members, intra-cluster percentage of migrant-households (proxy for migration networks), unemployed status of household head and access to additional financing have significant positive impact on migration, while coming from urban areas and household head's secondary education level has significant negative impact on migration, assuming that the family is relatively wealthy and might not need to send a member abroad to work and remit.

Yormirzoev (2017) empirically examines the factors influencing the labor migration flows from Tajikistan to Russia using World Bank 2007 Tajikistan Living Standard Survey (TLSS 2007) data. The study explores the key factors that drive labor migration from Tajikistan to Russia, focusing on both sociodemographic characteristics and migration-related social capital and finds that the demographic profile of migrants, particularly the relatively high share of younger individuals in the population, plays a significant role in shaping migration flows. Additionally, the knowledge of Russian as a second language is identified as a critical factor that facilitates migration to Russia.

The study by Sulaimanova et al. (2019) empirically investigates the drivers of international migration from Tajikistan and Kyrgyzstan using a gravity model over the period 1998–2011 and finds that economic factors such as GDP per capita, real wage(s), value added per worker in agriculture, remittances, exchange rates and demographic factor as amount of labor force significantly influence emigration decisions from Tajikistan and Kyrgyzstan.

4.3. Theoretical Framework and Variables

In choosing explanatory (X_i) and restricted (M_i) variables for our model, outlined in the next section, we rely on the theoretical models elaborated by Roy (1951), Borjas (1987, 1991), Clark, Hatton and Williamson (CHW, 2007) and Rapoport and Docquier (2005).

A) Variables determining the probability of having a migrant in a household

According to Borjas (1987, 1991) and CHW (2007) models, there are several motives that drive people to migrate.

- *Cost of Migration variables:*

The probability of migration rate is higher when relative migration costs are lower (Borjas, 1987). This can be reflected in the following relationship: as migration costs increase, then probability of migration should decrease.

$$\frac{\partial Pr(Migrate)}{\partial (Migration\ costs)} < 0 \qquad (4.1)$$

By CHW (2007) model, ***having social migrant networks*** in destination places can lower job search costs, the costs of finding affordable housing and childcare, and reduce vulnerability to exploitation, fraud, and crime. Also, having social networks with the same national, cultural origins can reduce the personal and cultural stresses associated with migration. In the context of Tajikistan this can be tested by including dummy variable that reflects if households have a network arranged in Russia.

Secondly, when ***members of a household can arrange the cost of migration*** (tickets, housing, food, fees, etc.) for one of its members – this significantly facilitates the migration process. Families might do so because of an "investment motive" – members of household might be interested in sending migrant to destination country because of expected high-returns to *all* family members/household (Rapoport, 2006). In the context of Tajikistan this can be tested by including dummy variable that reflects if households have arranged the cost of migration for their members.

Third, a number of studies revealed that having a ***legal status*** (Donato et al, 1992, 1993; Phillips et al, 1999; Massey et al, 2002, 2007) plays an important role in migration. If migrants do not have a legal status in their destination country they might receive lower wages, work longer hours, and are more likely to work in the informal economy. In the context of Russia, if foreign migrant plans to work in Russia they should obtain a legal status (work permit) in country (Russian Federation Parliament Decree № 1397, 2019). To obtain this work permit migrants should be contracted as workers by local company in Russia. And before hiring migrants, local companies test the skills level (job skills, language proficiency) of applicants and hire only the most skilled ones. So, we can consider a legal status of migrant as an implicit signal of knowledge and skills level, and also as a type of migration cost variable, since obtaining legal status before going to Russia significantly reduces the migration cost burden.

- *Education and skill variables:*

According to Roy's model of migration (1951), high-skilled and low-skilled workers would self-select themselves when they make choices to migrate to foreign country or not. The model predicts two possibilities. The first is *Positive Selection*: if high-skilled workers in the home country do not earn much more than the low-skilled workers (generally this occurs in developed countries), there will be positive selection in terms of migrants, so high-skilled workers will try to migrate to destination country that offer them higher wage premiums ("brain drain"). And low-skilled workers would prefer not to migrate. The second possibility is *Negative Selection*: If high-skilled workers could get high wages in their home country because of their knowledge and skills (generally this occurs in developing countries with high income disparity), then they would prefer not to migrate. But the low-skilled workers would prefer to migrate because of expected higher wage "insurance" provided by a narrower wage structure for their level of skills in the destination country.

In the context of Tajikistan, the dataset that is used in the analysis reveals that majority of workers that migrate mainly work in a ***low-skilled jobs***: construction (22%), agriculture (18%), trade (14%), cleaning, food preparation and sales, transport and communications (7%), utilities (6%), manufacturing (5%), etc. This actually reveals that we might have the classical Negative Selection case from above Roy's model for developing country: the low-skilled workers might prefer to migrate to Russia, because the expected wage for that level skills is higher in Russia, while high-skilled workers might prefer not to migrate because of decent payment for them in Tajikistan. The relationship between *knowledge and skills* in migration decisions of household members could be tested by including them into the selection equation. So, we also include several variables that reflect the ***knowledge and skills*** of household members.

We include a dummy variable which reflects ***if households have members with college level education***. According to number of previous studies it has been documented that higher education level might impact on migration decision of the person. In the context of Tajikistan, the dataset reveals that high school education might not be appropriate for testing the knowledge level, since majority of surveyed population (almost 80%) do have at least high school education (not enough variation). This motivates examination of college education level.

We include a proportion variable, which reflects the ***percentage of household members who have intermediate level of Russian language skills***. In the context of Tajikistan, learning Russian traditionally is more common in urban areas, while it is not so spread in rural areas (EurasiaNet, 2012).

B) Variables determining the amount of remittances transferred by migrants

To analyze the question of motives of sending the remittances we will mostly build upon Rapoport and Docquier work (2005), where authors gave theoretical basements for different types of motives to send remittances by migrants: altruism, exchange, strategic motive, insurance, moral hazard, family loan arrangement, inheritance, mixed motives.

In the case of Tajikistan, we think that the most appropriate motive for migrants to send remittances is *Exchange motive*, where migrants could be considered as "buying" or "paying" for the services that other household members do on their part in households in Tajikistan.

According to Rapoport and Docquier (2005), the evidence of Exchange motive of sending remittances might be the fact that *remittances increase with the quantity of "services" to be offered* – so family member (recipient) provides some "services" (taking care of family members, assets of migrant) and migrant is "paying" to recipient in exchange of those services. This can be reflected in the following relationship: as number of "services" increases then amounts of remittances also increases.

$$\frac{\partial(Remittances)}{\partial(No.of\ Services)} < 0 \qquad (4.2)$$

In the context of Tajikistan, it is common for the wife of the (male) migrant to take care of children and elderly family members. Therefore, the migrant might send more amounts of remittances to his home. We can proxy the variable of "services" with ***proportion of children and seniors in the household (percentage)***.

C) Other Household Demographic variables

Finally, we control for ***Household Size*** – number of members in the household. We also control for ***Household Income Beside Remittances*** – annual income of the household from all other sources and social allowances beside remittances. Finally, we control for a ***Wealth Index*** – the composite index constructed using Principal Component Analysis, which gives the scores based on ownership of different durable assets by households: number of rooms occupied by household, availability of individual heating system, water supply system and its status (operational or not), sewage system, status of power system (operational or not), availability of electric cooking appliances. In the context of Tajikistan, these variables reflect the wealth level of households, since after 1990's power supply, water supply, heating systems are outdated and had not been renovated systematically by the local governments because of lack of budget, and population try to renovate them privately (World Bank Report: Drinking Water, Sanitation and Hygiene Conditions in Tajikistan, 2017). More details are presented in Appendix-D4.

4.4. Empirical Model

The analysis is performed in two steps. First, the Probit estimation model is used to analyze the determinants of probability to have migrants in the households. The model for the probability of having a migrant in a household is estimated as follows:

$$Y_i = \alpha + \beta X_i + \gamma M_i + e_i \quad (4.3)$$

Y_i is a dichotomous variable which takes a value of 1 if household has at least one migrant. X_i is a vector containing explanatory variables. These include proxies for costs migration (networks, having legal status), education and skills (college degree and Russian language skills), proportion of children and seniors in household, and demographics (household size, household income beside remittances, wealth index).

Then, the Heckman two-step estimation method (1979) is used to analyze the determinants of probability to have migrants in the households and determinants influencing on remittances amount transferred. The Heckman method is used because remittances are observed only to subset of population (households that have migrants), so non-migrant households are not random and might have a self-selection bias. The model is estimated as follows:

$$Z_i = \alpha' + \delta X_i + u_i \qquad (4.4)$$

Z_i is an amount of remittances transferred. X_i is a vector containing all the explanatory variables defined in the following section. Error terms e_i and u_i follow a normal distribution $(0,1)$ and $N(0, \sigma_u)$, respectively and $Cov\ (e_i, u_i) = \rho$. X_i is a vector containing explanatory variables as in above.

To identify this model, we need to introduce an exclusion restriction: variables M_i which ensure that the error terms determining the probability of having migrants in a household and the amount of remittances transferred are not correlated. So, the selection equation (4.3.) has to contain at least one variable that does not exist in the equation (4.4) and thus does not explain the amount of remittances transferred. As M_i restriction variables we choose dummy variables if household has networks in destination place and if household has arranged the costs of migration for its members.

The Heckman estimation of correcting selection bias involves two steps: The first step is to estimate the probability of having migrants in the households (4.3). During this process an inverse Mill's ratio is calculated and will be included as an additional variable in the remittances equation (4.4).

This new term is $\lambda = \frac{\phi(\alpha+\beta X_i+\gamma M_i)}{\Phi(\alpha+\beta X_i+\gamma M_i)}$ (4.5)

Then we can re-write (3.4) as follows:

$$Z_i = \delta X_i + \rho\sigma_\varepsilon\lambda_i + \upsilon_i \qquad (4.6)$$

We can test for selection bias, because remittances can be observed only for migrants, so non-migrants are not random. We need to test if there is an omitted variable: test if coefficient of λ_i equals zero under H_0 (H_0: $\rho = 0$), which means that errors from two equations are not correlated. The existence of a selection bias is tested by the assumption that the estimated coefficient of the inverse Mills ratio is statistically significant. In this case, the two-step method allows obtaining unbiased estimates of the coefficients.

4.5. Data

The data used in this study came from cross-sectional Tajikistan World Bank Microdata "Jobs, Skills, and Migration Survey" for 2013. It provides information on demographic and socio-economic characteristics of households and their members. It also provides information on the existence of migrant family members and detailed information on the characteristics of these members. This survey is conducted by a questionnaire developed by the World Bank. The sampling procedure has ensured that the sample is highly representative. It generates a large country survey dataset on more than 24,000 individuals and 3300 households. Table-4.1 presents data on the main characteristics of the households surveyed according to the receipt or not of international remittances, demographics and migration information. The analysis is done using household level data.

Table-4.1. Summary Statistics

Variable	Observations	Mean	Std. Deviation	Min	Max

Household Size	3,300	6.10	3.07	1	19
Ave. Remittances (Tajik Somoni)	1,188	2,121.16	4,139.71	10	70000
Proportion of Children (%)	3,296	34.97	21.39	0	100
Proportion of Seniors (%)	3,300	8.20	17.18	0	100
Proportion of HH Members with Good Russian Skills (%)	3,299	57.07	33.76	0	100
Proportion of HH Members with Secondary Education (%)	2,606	43.25	22.67	6	100
Proportion of HH Members with College Education (%)	3,300	13.54	21.93	0	100
D. If HH has a Member with College Education (Y=1, N=0)	3,300	0.407	0.491	0	1
D. If Networks Arranged (Y=1, N=0)	3,300	0.367	0.482	0	1
D. If HH had Migrant last year (Y=1, N=0)	3,300	0.307	0.461	0	1
D. If HH Members have Legal Status to Work in Russia (Y=1, N=0)	3,300	0.311	0.463	0	1
D. If HH has Arranged Costs of Migration (Y=1, N=0)	3,300	0.085	0.280	0	1
Wealth Index category	3,300	2.339	1.733	1	5
Annual Income of HH beside Remittances (Tajik Somoni)	2,975	1,593.47	8,448.66	10	299,000

We note firstly that out of 3,300 Tajik households, 1,188 households receive remittances. The study revealed that the proportion of children in households is 34% on average, and proportion of seniors in the households is around 17% on average. Proportion of household members with secondary education account for 43% on average, with college education – for 13%[8] on average.

[8] Tajikistan has seen a notable rise in university enrollment since 1991, with gross enrollment ratio for tertiary education of 31.1% in 2017, but still lower than in many other post-Soviet states (Wetzinger, 2024).

We have number of dummy variables that account for if members have college education, if households arrange migration costs for its members, if networks are arranged with those households in destination place, if members have legal status to work in Russia and if households have migrant members. Also, households are sub-divided into 5 categories according to their wealth level: poor, medium low, medium, medium upper, rich households. We can note that on average households fall into the medium-low wealth category. The annual income of households besides remittances is also reported from survey, with a mean of 1,593 Tajik Somonis. A correlation matrix is presented in *Appendix-D1*.

4.6. Results and Discussion

4.6.1. Marginal Effects of Probit Model

Probit regression coefficients for determining the probability of having a migrant in a household are reported in Table-D2 in the *Appendix-D* for reference.

To understand the model better we calculate the marginal predicted probabilities of having a migrant in households at average level of particular predictor variables, holding all other variables in the model at their means. Results are presented for "All Households", "Urban Households" and "Rural Households" subsets in Table-4.2.

Table-4.2. Marginal Effects of Probit Regression Analysis of Households

	(Probit: All)	**(Probit: Urban)**	**(Probit: Rural)**
VARIABLES	**Prob. of HH to have Migrant (Y=1, N=0)**	**Prob. of HH to have Migrant (Y=1, N=0)**	**Prob. of HH to have Migrant (Y=1, N=0)**
HH Size	0.0223***	0.0194***	0.0205***
	(0.00316)	(0.00357)	(0.00524)
Proportion of Children (%)	-0.00259***	-0.00209***	-0.00242***
	(0.000468)	(0.000480)	(0.000858)
Proportion of Seniors (%)	-0.00133**	-0.00124**	-0.00103

	(0.000619)	(0.000596)	(0.00135)
1D. If HH has a Member with College Education	-0.0806***	-0.0436**	-0.0620
	(0.0183)	(0.0188)	(0.0381)
Proportion of HH Members with Intermediate Russian Skills (%)	-0.000778**	-7.98e-05	-9.33e-05
	(0.000305)	(0.000339)	(0.000575)
Annual Income of HH (without Remittances) (Tajik Somoni)	7.65e-07	-1.17e-07	1.26e-06
	(1.02e-06)	(9.40e-07)	(1.78e-06)
Wealth Index categories:			
2. Wealth Index (below medium)	0.00886	-0.0261	0.0524
	(0.0288)	(0.0259)	(0.111)
3. Wealth Index (medium)	-0.0220	-0.0813***	-0.0486
	(0.0279)	(0.0252)	(0.106)
4. Wealth Index (upper medium)	0.0143	-0.00994	-0.0475
	(0.0303)	(0.0322)	(0.107)
5. Wealth Index (rich)	0.0190	-0.0490	-0.0274
	(0.0300)	(0.0301)	(0.106)
1D. If Networks Arranged	0.279***	0.258***	0.267***
	(0.0197)	(0.0260)	(0.0303)
1D. If HH has Arranged Costs of Migration	0.356***	0.387***	0.314***
	(0.0388)	(0.0615)	(0.0466)
Observations	2,970	1,814	1,156

*Note: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.*

The following results can be seen from the marginal effects:

Wealth Index: Overall, the results demonstrate that propensity to migrate is "u-shaped" with household's income: The significant negative magnitude is demonstrated by "Urban" subset of households for Wealth-3 category (medium income): the predicted probability of having a migrant in household decreases by 8 percentage points, holding other variables at their means. The result demonstrates that being from middle income households in urban areas decreases the probability of having a migrant, probably because of more favorable economic situations of medium income families in urban areas and wealth level in general. The results for medium income level households for "Rural" and "All" subsets demonstrate the same negative direction, even though they remain as not significant, indicating that families with more favorable economic situation in general might tend to migrate less, no matter of their location in urban or rural zones.

Meanwhile, the households from low-income category demonstrate interesting results: those, who are located in urban areas, might tend to migrate less (negative direction, even not significant), but in general all households in the subset for that group and especially those in rural areas - demonstrate that might tend to migrate more (positive direction, even not significant). So, it shows that due to more harsh economic situation in rural areas and overall because of less income, people tend to migrate more. It aligns with the fact from literature that many Tajiks migrate to Russia due to limited job prospects and lower salaries (which reflects low-income households' population) in Tajikistan compared to Russia (Olimova, 2010).

The results from high-income households' groups indicate (even though not significantly) that overall they tend to migrate more, but if we look on rural/urban breakdown, it demonstrates an opposite negative direction - tend to migrate less (even though not significantly). This might be because high-income households have more opportunities to migrate – they can afford it without too much economic obstacles, and get higher-paying jobs or start their own small-businesses in Russia, which benefits their families better back in Tajikistan. This aligns with existing literature reporting about brain drain of high-skilled (proxied possibly by high-income households) people from Tajikistan to Russia (Vasilyevich et al, 2024).

Family size: This variable is positively significant for all subsets ("All", "Urban" and "Rural"). For "All" subset of households, as there are more family members, the probability of having a migrant increases by 2.23 percentage points, holding other variables at their means. When comparing urban and rural areas, the stronger association is revealed in rural areas than in urban areas: more family members increase the predicted probability of having a migrant in a household by 2.05 percentage points in rural areas and by 1.94 percentage points in urban areas, holding other variables at their means. This happens probably because of more difficult economic situation and less career prospects in rural areas in general, so people try to migrate to other countries.

***Proportion of Children*:** This variable is negatively significant for all subsets ("All", "Urban" and "Rural"). For "All" subset of households, as there are more children, the probability of having a migrant decreases by 0.25 percentage points, holding other variables at their means. Comparing urban and rural areas shows that, in rural areas the association is stronger than in urban areas: as there are more children in a household, the predicted probability of having a migrant decreases by 0.24 percentage points in rural areas and by 0.20 percentage points in urban areas, holding other variables at their means. This happens probably because of less alternative opportunities of caring after children and more limited financial resources of households in rural areas, so family members have to stay in Tajikistan to take care of their children.

Proportion of Seniors: This variable is negatively significant for two subsets ("All" and "Urban"). For "All" subset of households, as there are more seniors, the probability of having a migrant decreases by 0.13 percentage points, holding other variables at their means. In "Urban" subset of households: as there are more seniors in a household, the predicted probability of having a migrant decreases by 0.12 percentage points, holding other variables at their means. For "Rural" subset this variable remains as not significant. The result suggests that for households in all subsets and urban areas the probability of having a migrant is decreasing with more seniors in the household, probably because of age limitations and cultural traditions of caring for seniors by other members of the household.

***Having College Degree**:* This variable is negatively significant for two subsets ("All" and "Urban"). For "All" subset of households, having a college education decreases the probability of having migrant by 8.06 percentage points, holding other variables at their means. In "Urban" subset of households, having a college education decreases the probability of having migrant by 4.36 percentage points, holding other variables at their means. For "Rural" subset this variable remains as not significant. This happens probably because of better job prospects in urban areas in Tajikistan due to higher level of education.

Knowing Russian Language: This variable is negatively significant for "All" subset of households. In "All" subset of households: as there are more members who speak Russian language in a household, the predicted probability of having a migrant decreases by 0.07 percentage points, holding other variables at their means. This is probably because of better job prospects in local areas in Tajikistan due to speaking of Russian language, which demonstrates higher educational skills. *Note:* At present time, even the state language is Tajik, but the 1994 Constitution of the Republic of Tajikistan (Article-2) recognizes Russian language as a means of inter-ethnic communication and allows its use in some cases for official paperwork (Saidov et al, 2020). For "Urban" and "Rural" subsets this variable remains as not significant.

Arranged Cost of Migration: This variable is positively significant for all subsets ("All", "Urban" and "Rural"). For "All" subset of households, arranging the cost of migration increases the probability of having migrant by 35.6 percentage points, holding other variables at their means. When comparing urban and rural areas, in urban areas the association is stronger than in rural areas: the predicted probability of having a migrant increases by 38.7 percentage points in the households with arranged costs of migration in urban areas and by 31.4 percentage points in rural areas, holding other variables at their means. It probably happens because of more favorable financial status of households in urban areas and more limited financial resources of households and general less favorable economic situation in rural areas.

Arranged Networks: This variable is positively significant for all subsets ("All", "Urban" and "Rural"). For "All" subset of households, arranging the cost of migration increases the probability of having migrant by 27.9 percentage points, holding other variables at their means. When comparing urban and rural areas, in rural areas the association is stronger than in urban areas: arranging the cost of migration increases the probability of having migrants by 26.7 percentage points and by 25.8 percentage points in urban areas, holding other variables at their means. This is probably because of stronger family-friends ties in rural areas than in urban areas and also probably because of less desirable economic situation in rural areas with less available jobs, so people need to migrate more to other places to find a job.

In general, by analyzing marginal predicted probabilities for different variables, holding other variables at their means for different subsets, we can see that majority of results for urban and rural areas demonstrate similar behavior. That is why "All" subset of households is chosen for further analysis of Heckman two-step estimation.

4.6.2. Sensitivity Analysis: Marginal Effects of Probit Model, using Consumption Quantiles

In previous section we see results of marginal effects from Probit Model when we used constructed Wealth Index variable (based on household living conditions) as some sort of proxy that reveals the *real* economic situation of the households. To ensure that results are stable, we aimed to check those marginal effects by estimating Probit Model by using *Consumption Quantiles* variable instead of *Wealth Index*, as a proxy for real economic situation of the households.

Consumption Quantile Index variable is already constructed in the Tajikistan World Bank dataset based on per capita real consumption behaviors of the households in Tajikistan for 2013 *(World Bank-Tajikistan Jobs, Skills, and Migration Survey, 2013).*

As a robustness check, Probit regression coefficients for determining the probability of having a migrant in a household, but with using *Consumption Quantiles* instead of *Wealth Index*, are reported in Table-D3 in the *Appendix-D* for reference.

To understand the model better we calculate the marginal predicted probabilities of having a migrant in households at average level of particular predictor variables, holding all other variables in the model at their means. Results are presented for "All Households", "Urban Households" and "Rural Households" subsets in Table-4.3.

Table-4.3. Marginal Effects of Probit Regression Analysis of Households, using Consumption Quantiles

	(Probit: All)	**(Probit: Urban)**	**(Probit: Rural)**
VARIABLES	**Prob. of HH to have Migrant (Y=1, N=0)**	**Prob. of HH to have Migrant (Y=1, N=0)**	**Prob. of HH to have Migrant (Y=1, N=0)**
HH Size	0.0206***	0.0143***	0.0177***
	(0.00311)	(0.00353)	(0.00525)
Proportion of Children (%)	-0.00269***	-0.00215***	-0.00246***
	(0.000470)	(0.000484)	(0.000863)
Proportion of Seniors (%)	-0.00139**	-0.00142**	-0.00107
	(0.000621)	(0.000603)	(0.00135)
1D. If HH has a Member with College Education	-0.0748***	-0.0289	-0.0631*
	(0.0184)	(0.0191)	(0.0381)
Proportion of HH Members with Intermediate Russian Skills (%)	-0.000747**	-1.54e-05	0.000106
	(0.000302)	(0.000340)	(0.000569)
Annual Income of HH (without Remittances) (Tajik Somoni)	7.35e-07	3.20e-08	1.37e-06
	(1.03e-06)	(9.54e-07)	(1.82e-06)
Consumption Quantiles			
2. Quintile_2013_TJ (below medium)	-0.00858	-0.0116	-0.0335
	(0.0297)	(0.0336)	(0.0529)
3. Quintile_2013_TJ (medium)	-0.0455	-0.0806***	-0.0420
	(0.0286)	(0.0311)	(0.0512)
4. Quintile_2013_TJ (upper medium)	-0.0130	-0.0530*	-0.0195
	(0.0289)	(0.0319)	(0.0509)
5. Quintile_2013_TJ (rich)	-0.0573**	-0.0697**	-0.0827

	(0.0275)	(0.0299)	(0.0512)
1D. If Networks Arranged	0.283***	0.257***	0.268***
	(0.0195)	(0.0259)	(0.0302)
1D. If HH has Arranged Costs of Migration	0.352***	0.363***	0.310***
	(0.0389)	(0.0612)	(0.0468)
Observations	2,970	1,814	1,156

*Note: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.*

The presented results show that these marginal effects are in general similar to the marginal effects from Probit Model using ***Wealth Index*** presented in the previous part of the paper, with a slight difference in magnitudes.

Family size and Arranged Networks variables still have a positive relationship to probability of migration, stronger for households in rural areas, probably because of more difficult economic situation and less career prospects in rural areas in general.

Arranged Cost of Migration variable also has a positive relationship to probability of migration as before, stronger for households in urban areas, rather than rural areas: probably because of more favorable financial resources of households in urban areas in general.

Having Dependents (children and seniors) has negative relationship to probability of migrating due to caring responsibilities for “All” subset, similar as before.

Education skills (College education and Russian language skills) still demonstrate negative relationship to probability of migration as before: more skilled people tend to find jobs locally in Tajikistan than migrating. The difference with previous marginal effects is that College education variable is significant for “All” and “Rural” subsets, instead of “All” and “Urban” subsets as in previous results.

Another difference from previous model results is in ***Consumption Quantiles*** (which is used instead of *Wealth Index*): similar to previous results, the significant negative magnitude is demonstrated by "Urban" subset of households for Quantile-3 category (medium income), similar to before. Additionally, the significant negative results are demonstrated by “All” and "Urban" subsets of households for Quantile-5 category (high income). So, expenditure quantiles are negatively significant for households in quantiles 3, 5 in “All” and “Urban” subsets: the higher are consumptions in households (richer), the less they tend to migrate, compared to poorest households.

In general, by analyzing the marginal predicted probabilities using ***Consumption Quantiles***, we see that results are similar to marginal effects from using *Wealth Index*, which suggests that sensitivity analysis produces similar results as before, and that results are firmly stable, even different household income proxy variables are used.

4.6.3. Heckman Model Results

The Heckman method is used because remittances are observed only to subset of population (households that have migrants), so non-migrant households are not random. To address this self-selection problem and to get unbiased results for both remittances equation and probability of having migrants equation, the two-step Heckman (1979) estimation is performed and regression results are reported in Table-4.4.

Table-4.4. Heckman Regression Analysis of Households: migrants and remittances transferred

	(1)	(2)	(3)
VARIABLES	**Remittances (Tajik Somoni)**	**Prob. of HH to have Migrant (Y=1, N=0)**	**Mills Ratio**
HH Size	109.4**	0.0742***	

	(46.43)	(0.0124)	
Proportion of Children (%)	2.807	-0.0103***	
	(8.282)	(0.00200)	
Proportion of Seniors (%)	-16.04	-0.00626**	
	(14.06)	(0.00305)	
D. If HH has a Member with College Education (Y=1, N=0)	-44.85	-0.179**	
	(322.6)	(0.0810)	
Proportion of HH Members with Intermediate Russian Skills (%)	8.744*	-0.00238*	
	(5.202)	(0.00133)	
Annual Income of HH (Tajik Somoni)	0.0146*	3.15e-06	
	(0.00751)	(3.58e-06)	
1.Wealth Index (poor)	100.6	-0.116	
	(530.1)	(0.128)	
2.Wealth Index (below medium)	-13.93	-0.195	
	(569.7)	(0.132)	
3.Wealth Index (medium)	-110.1	-0.272**	
	(426.1)	(0.113)	
4.Wealth Index (upper medium)	-152.4	-0.104	
	(414.9)	(0.118)	
5.Wealth Index (rich)	221.5	-0.0560	
	(387.0)	(0.108)	
D. If HH Members have Legal Status to Work in Russia (Y=1, N=0)	-857.2**	0.572***	
	(361.0)	(0.0888)	
D. If HH has Arranged Costs of Migration (Y=1, N=0)		1.026***	
		(0.113)	
D. If Networks Arranged (Y=1, N=0)		0.920***	
		(0.0919)	
Lambda			-801.8**
			(373.7)
Constant	1,512*	-1.511***	
	(777.7)	(0.164)	
Observations	2,586	2,586	2,586

*Note: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.*
Source: Authors calculations.

In the output: lambda, which is the coefficient on the Inverse Mills Ratio, represents the covariance between the errors in the remittances and the selection equation under the assumptions of the model. The parameter ρ is the correlation between these error terms. In our model, lambda is estimated to be statistically significantly negative (-801.8), which implies that there is a significant selection bias (errors from two equations are correlated, reject null) and therefore we should use the Heckman estimation method to correct for self-selection bias, which gives us unbiased coefficient estimates. It can be seen from Table-4 that several household and migrant variables significantly affect the likelihood of having migrants in the households and the amount of remittances transferred.

- **Variables determining the probability of having a migrant in a household**

The results in Table-4.4 show that migration cost variables like *having networks* in destination places and ***family arrangement of costs*** related to migration of its members have significant positive relationship with probability of having migrants in the households. This result proves our initial hypothesis based on Borjas (1987, 1991) and CHW (2007) models, that when migration costs decrease then migration rate might increase.

As it was discussed earlier, ***having networks*** in Russia with the same national, cultural and traditional background significantly reduces the cost of migrating and adapting in a new place. Tajik migrants find the jobs in Russia mostly before departing from Tajikistan through these types of social networks. While working in Russia, social networks help migrants in number of daily basic needs: how and where to find affordable housing, where to ask for legal assistance if needed, what are the most efficient travel routs, how to get medical insurance and where to get medical assistance if needed, etc.

The *arrangement of migration costs for migrant by family* could reveal the "investment motive" of the whole family. In context of Tajikistan, where several close families traditionally live in one household (ex: senior parents, two brothers with wives and children) – sending usually male members to Russia for work is a strategic plan for the whole family because of a job availability and expected higher wages in Russia, which is sufficient to maintain the whole household. This happens because Tajikistan unfortunately still experiences the lack of job places and adequate payment which could be enough for living in many industries.

Table-4.4 reports that ***college education*** and ***knowing Russian language*** has statistically significant negative association with the probability of having migrants in the households. This result actually supports our initial hypothesis about Negative Selection in Tajikistan that is predicted by Roy's model (1951). As it is discussed earlier, the analyzed dataset reveals that majority of migrants who migrate to Russia generally work in low-skill jobs, so this result proves that the low-skilled workers might prefer to migrate to Russia, because the expected wage for that level skills is higher in Russia, while high-skilled workers, who have college level education and more likely to know Russian language, might prefer not to migrate because of decent payment for them in Tajikistan, especially in urban areas.

On contrary, having ***legal status*** before departing to Russia has a significant positive association with probability of having migrants in the households. This might happen because of dual nature of this variable – it is considered both as migration cost variable and implicit skills variable. The positive correlation of legal status reveals more of migration cost nature: since migrants have to pass a lot of different tests before obtaining this status, it means that their migration cost burdens are decreased once they obtain legal status, so migration rate might increase.

Having dependents in the household (*children and seniors*) has significant negative association with migration decisions. This result matches the cultural traditions in Tajikistan, where if family has little children or senior grandparents, they prefer to stay in Tajikistan and take care of their dependents. But overall *family size* has significant positive association with migration decision – this is intuitive, the bigger is a family, the higher is possibility to have migrants because of need to sustain the whole family and find better job opportunities in Russia, rather in Tajikistan.

Household wealth level has significant association with migration decision only in medium-income level households: it shows that medium income families are less prone to send migrants to Russia, because they might have other sufficient income sources in Tajikistan.

- ***Variables determining the amount of remittances transferred by migrants***

The results in Table-4.4 show that ***family size*** and ***family income*** have significant positive relationship with remittances transferred to households. That matches with intuitional predictions: the bigger is a family the more remittances transfer to be expected - to take care of all family members; and the richer is a family, then the more remittances could be transferred – because family could afford to send more migrants to Russia and expect higher returns from remittances.

It should be noted that ***proportions of dependents*** in the household (*children and elderly*) seem not to have a relationship with the amount of remittances, as we initially hypothesized based on Rapoport and Docquier work (2005).

Knowing Russian language has significant positive relationship with amount of remittances – since it would be easier to find higher-paid job knowing the language in Russia, and so more remittances could be sent to Tajikistan.

Having legal status to work in Russia seems to have significant negative relationship with the amount of remittances transferred to Tajikistan – this is counter-intuitive outcome, which might be observed because we analyzed the data from 2013, when the migration legislation in Russia for citizens from Tajikistan has not been yet improved (2019) and the cost of obtaining legal work permit was so high and frequent, so it significantly reduced the amounts of remittances sent by migrants to Tajikistan.

The Heckman two-step analysis also was performed on "Urban" and "Rural" subsets. Results are available from the author upon request, though we note that the smaller sample sizes for this disaggregation reduced the statistical power of the tests.

4.7. Conclusions

The paper examines the motives of migration from Tajikistan and sending remittances to households in Tajikistan based on previous theoretical models. The World Bank Tajikistan Migration dataset for 2013 is used to analyze those motives. The first research question analyzed in the paper is why people migrate from Tajikistan. To answer this, the Probit Model is used by examining migrant and household characteristics as determinants of migration from Tajikistan. The second research question is to understand why migrants remit to households in Tajikistan. To answer this question two-step Heckman estimation is performed by using migrant and household characteristics to address for self-selection bias during analyzing the motives of migration and sending remittances.

The marginal effects from the Probit model, determining the probability of having migrants in the households for different subsets, reveal that being from *middle income* households in urban areas has a negative relationship to the probability of having a migrant, probably because of more favorable economic situation in urban areas and wealth level in general: the predicted probability of having a migrant in household decreases by 8 percentage points, holding other variables at their means. In “All” and “Rural” subsets for *middle income* group and for households from *upper-income* level this result demonstrates opposite positive direction (even though not significant). Overall, the results demonstrate that propensity to migrate is “u-shaped” with household’s income: while people from middle-income group tend to migrate less due to more stable economic position in their households, people from low income groups seem to be pushed to migrate due to economic reasons - mainly because of lack of job opportunities in Tajikistan, but also people from rich households also seem to migrate - probably because their high skills could possible make it easier for them to find high-paid jobs in destination country, reflecting “brain drain” process of exodus of high-skilled professionals from Tajikistan.

Family size has a positive correlation with the probability of having a migrant and this is stronger for households in rural areas, rather than urban areas, probably because of more difficult economic situation and less career prospects in rural areas, in general: additional family members increase the predicted probability of having a migrant by 2.05 percentage points in rural households and by 1.94 percentage points in urban households, holding other variables at their means.

Having Dependents (*Children,* Seniors) has a negative association with the probability of migrants, probably because of more caring responsibilities, so adults have more limited factors in their migration decisions. Also, the results show that people from rural areas tend to migrate less than people from Urban areas if they have more dependents: probably because of less alternative opportunities for caring after dependents, more stronger family ties and less favorable economic situation in rural areas in general: as there are more children in a household, the predicted probability of having a migrant decreases by 0.24 percentage points in rural areas and by 0.20 percentage points in urban areas, holding other variables at their means; as there are more seniors in a household, the predicted probability of having a migrant decreases by 0.12 percentage points in urban areas, holding other variables at their means and being not significant in rural areas.

Having a College Degree and Knowing Russian language have a negative relationship with the probability of having migrant in “All” and “Urban” subsets of households, probably because of more possibilities to find a job locally in urban areas due to higher educational skill levels: the predicted probability of having a migrant in a household decreases by 4.36 percentage points for the households with members that have College Education in urban areas, holding other variables at their means and being not significant in rural areas. The more household members in “All” subset of households know Russian language, the predicted probability of having a migrant decreases by 0.07 percentage points, holding other variables at their means.

Arranged Cost of Migration and Networks have a positive association with the probability of having migrants in all subsets, so it is easier to migrate since the whole family finances/arranges the migration cost for its member. Also, the results show that, on one hand, people from urban areas tend to migrate more than people from rural areas because of *covered/arranged costs of migration* by families: the predicted probability of having a migrant increases by 38.7 percentage points in the households with arranged costs of migration in urban areas and by 31.4 percentage points in rural areas, holding other variables at their means, probably because of more favorable financial status in urban areas. On the other hand, people from rural areas tend to migrate more than people from urban areas because of more *arranged networks* and less favorable economic situation in rural areas in general: the predicted probability of having a migrant increases by 26.7 percentage points for the households with arranged networks of migration in rural areas and by 25.8 percentage points in urban areas, holding other variables at their means, probably because of more stronger family-friends ties in rural areas than in urban areas and also probably because of less desirable economic situation in rural areas, with less available jobs.

As a robustness check, the Probit regression coefficients for determining the probability of having a migrant in a household, but with using *Consumption Quantiles* instead of *Wealth Index* are analyzed and reveal similar results as before.

The empirical results from two-step Heckman estimation show that migration cost variables like *having networks* in destination places and family *arrangement of costs* related to migration of its members have positive correlation with the probability of having migrants in the households. This result matches the prediction of Borjas (1987, 1991) and CHW (2007) models, that when migration costs decrease then migration rate might increase. *The education and skills variables* reveal negative association with a probability of having migrants in the households, which supports *Negative Selection* migration process predicted Roy's model (1951), that is going on in Tajikistan. *Language skills* showed a positive association with the amount of remittances, since it would be easier to find higher-paid job knowing the language in Russia, and so more remittances could be sent to Tajikistan. *Family size* is revealed to have a positive relationship with migration decisions, while *wealth level* demonstrates a negative relationship but only for medium-income level households. Also, *family size and family income* show a positive relationship with the remittances transferred to households. Additionally, the results reveal that there is a significant selection bias (errors from two equations are correlated, lambda is negative and significant) and therefore Heckman estimation method use is justified to correct for the self-selection bias.

As a policy implication, this paper empirically reinforces the evidence of how important migration networks and family are arranged costs of migration in shaping the migrations flows from Tajikistan to Russia, making diasporas important “non-formal” institutions between two countries, since they result in the highest positive magnitudes in migration decisions in the households. Also, contrary to the well-accepted theory that education and skills push the most talented people to migration, in the context of Tajikistan we actually see the opposite, Negative Selection, when more unskilled and less educated people tend to migrate more, especially from rural areas. Additionally, results show that people from rural areas tend to migrate more, which describes how still weak is the economic situation in Tajikistan, especially in rural areas, with high levels of unemployment and less career prospects for local people, so they have to migrate to provide substance support.

REFERENCE:

1. Abdullaev, I., Kazbekov, J., & Rasulov, A. (2015). Migration and the economy in Tajikistan: A case study of Central Asia. *Migration Policy Institute*.
2. Abdulloev, I., Gang, I. N., & Landon-Lane, J. (2012). Chapter 6 Migration as a substitute for informal activities: Evidence from Tajikistan. In H. Lehmann & K. Tatsiramos (Eds.), *Informal employment in emerging and transition economies* (Research in Labor Economics, Vol. 34, pp. 205–227). Emerald Group Publishing Limited. https://doi.org/10.1108/S0147-9121(2012)0000034009
3. Acharya, C. P., & Leon-Gonzalez, R. (2014). How do migration and remittances affect human capital investment? The effects of relaxing information and liquidity constraints. *Journal of Development Studies, 50*(4), 444-460. https://doi.org/10.1080/00220388.2013.866224
4. Adams, R. H., & Cuecuecha, A. (2010). Remittances, household expenditure and investment in Guatemala. *World Development, 38*(11), 1626–1641.
5. Amare, M., Abay, K., Arndt, C., & Shiferaw, B. (2021). Youth Migration Decisions in Sub-Saharan Africa: Satellite-Based Empirical Evidence from Nigeria. Population and Development Review. https://doi.org/10.1111/padr.12383
6. Amuedo-Dorantes, C. (2023). The widespread impacts of remittance flows. *IZA World of Labor*. Retrieved from https://wol.iza.org
7. Asian Development Bank (ADB). (2019). Tajikistan: Economy overview and labor force statistics.
8. Balli, F., & Rana, F. (2015). Determinants of risk sharing through remittances. *Journal of Banking & Finance, 55*, 107-116. https://doi.org/10.1016/j.jbankfin.2015.02.003
9. Borjas, G. J. (1987). Self-selection and the earnings of immigrants. *American Economic Review, 77*(4), 531–553. https://www.jstor.org/stable/1814529
10. Borjas, G. J. (1991). Immigration and self-selection. In J. M. Abowd & R. B. Freeman (Eds.), *Immigration, trade, and the labor market* (pp. 29–76). University of Chicago Press. https://www.nber.org/chapters/c6663
11. Borzhikova, A. (2014). *Demographic report on Tajikistan's population structure*. State Statistical Agency.
12. Carling, J. (2008). The determinants of migrant remittances. *Oxford Review of Economic Policy,* 24(3), 581–598. https://doi.org/10.1093/oxrep/grn022
13. Williamson, J. G., Clark, X., & Hatton, T. J. (2007). Explaining U.S. immigration, 1971-1998. Review of Economics and Statistics, 89(2), 359–373. https://doi.org/10.1162/rest.89.2.359
14. Contreras, S. (2012). The influence of migration on human capital development. Asian and Pacific Migration Journal, 21(3), 365-384. https://doi.org/10.1080/10168737.2012.659277
15. Czaika, M., & Vothknecht, M. (2014). Migration and aspirations – are migrants trapped on a hedonic treadmill? IZA Journal of Migration, 3, Article 1. https://doi.org/10.1186/2193-9039-3-1.
16. Dao, T. H., Docquier, F., Parsons, C., & Peri, G. (2018). Migration and development: Dissecting the anatomy of the mobility transition. *Journal of* Development Economics, 132, 88–101.https://doi.org/10.1016/j.jdeveco.2017.12.003
17. Dinkelman, T., & Mariotti, M. (2016). The long-run effects of labor migration on human capital formation in communities of origin. *American Economic Journal: Applied Economics, 8*(4), 1–35. https://doi.org/10.1257/app.20150405

18. Dustmann, C., & Mestres, J. (2010). Remittances and temporary migration. *Journal of Development Economics, 92*(1), 62–70.
19. Filmer, D., & Pritchett, L. (2001). Estimating wealth effects without expenditure data—or tears: an application to educational enrollments in states of India. *Demography*, 38(1), 115-132. https://doi.org/10.1353/dem.2001.0003
20. Chami, R., Hakura, D., & Montiel, P. J. (2009). Remittances: An automatic output stabilizer? *IMF Working Paper No. 09/91*. Retrieved from https://ssrn.com/abstract=1394811
21. Goldin, I., Cameron, G., & Balarajan, M. (2011). *Exceptional people: How migration shaped our world and will define our future*. Princeton University Press.
22. Gorbachev, S. (2017). Temporary labor migration from Tajikistan to Russia. *Journal of Ethnicand Migration Studies, 43*(10), 1783–1801.
23. *Grogger, J., & Hanson, G. H. (2011). Income maximization and the selection and sorting of international migrants.* Journal of Development Economics*, 95(1), 42–57.* https://doi.org/10.1016/j.jdeveco.2010.05.001
24. *Heckman, J. (1979). Sample selection bias as a specification error. Econometrica, 47(1), 153–161.* https://doi.org/10.2307/1912352
25. Japan International Cooperation Agency (JICA). (2018, 2019). Migration survey report on Tajikistan. https://www.jica.go.jp/english/jica_ri/news/topics/2019/20190220_01.html
26. Kuznetsov, Y. (2006). *Diaspora networks and the international migration of skills: How countries can draw on their talent abroad*. World Bank Publications. Retrieved from https://books.google.com
27. Labonne, M., Biller, A., & Chase, M. (2007). Measuring the multidimensionality of poverty: An application to Vietnam. *Social Indicators Research*, 81(3), 535-562.
28. Lee, P. (2023). *Challenges of migration and remittance dependence in Tajikistan*. *Global Migration Review*.
29. Liu, M. (2013). *Migrant networks and international migration: Testing weak ties.* Demography 50: 1243–1277. https://doi.org/10.1007/s13524-013-0213-5
30. Luecke, M., Omar Mahmoud, T., & Steinmayr, A. (2013). Labor migration and remittances in Tajikistan. *Working Paper*.
31. Mahat, I., Schuettler, K., & Sijapati, B. (2017). Gender, migration, and remittances in Tajikistan. *Asian Development Bank Report*.
32. Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., & Taylor, J. E. (2013). *Worlds in motion: Understanding international migration at the end of the millennium*. Oxford University Press.
33. Meier, A. (2014). *Determinants of low-skilled labor migration in Tajikistan*. World Bank Tajikistan Living Standards Measurement Survey. https://doi.org/10.2139/ssrn.2550701
34. McKenzie, D. (2005). Principal component analysis for creating a wealth index. *Caltech Economics Working Paper*, (05-001).
35. Minujin, A., & Bang, H. (2002). Measuring household wealth: A proposed methodology. *The World Bank Economic Review*, 16(1), 35-62.
36. Naudé, W. (2010). The Determinants of Migration from Sub-Saharan African Countries, *Journal of African Economies*, Volume 19, Issue 3, June 2010, Pages 330–356, https://doi.org/10.1093/jae/ejq004
37. Olimova, S. (2010). The Impact of Labour Migration on Human Capital: The Case of Tajikistan. *Revue européenne des migrations internationales*, vol. 26 - n°3. https://doi.org/10.4000/remi.5239

38. Rapoport, H., & Docquier, F. (2005). The economics of migrants' remittances. *IZA Discussion Paper No. 1531*. https://doi.org/10.1016/S1574-0714(06)02017-3
39. Ratha, D., et al. (2022). *Remittances and their socio-economic effects on Tajikistan*. World Bank.
40. Roberts, K., & Jones, R. (2020). *Post-Soviet migration trends in Tajikistan*. *Migration Studies*.
41. Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers, 3*(2), 135–146. https://doi.org/10.1093/oxfordjournals.oep.a041827
42. Ruyssen, I., & Rayp, G. (2014). Determinants of Intraregional Migration in Sub-Saharan Africa 1980-2000. *The Journal of Development Studies*, *50*(3), 426–443. https://doi.org/10.1080/00220388.2013.866218
43. Schwartz, M. (2014). Tajik migrants and the Russian labor market. *Central Asia Program Paper No. 151*.
44. Siddiqui, T., & Turton, R. (2008). Migrant workers' remittances and micro-enterprise development: Bangladesh case study. *ILO Report*.
45. Sulaimanova, B., & Bostan, A. (2014). International migration: A panel data analysis of the determinants of emigration from Tajikistan and Kyrgyzstan. *Eurasian Journal of Business and Economics,* *7*(13), 1-9. https://www.ejbe.org/EJBE2014Vol07No13p001SULAIMANOVA-BOSTAN.pdf
46. Tajikistan Statistics Agency. (2013). *Macroeconomic Indicators of Tajikistan*. http://old.stat.tj/en/macroeconomic-indicators
47. United Nations Development Programme (UNDP). (2014). *Tajikistan Human Development Report: Employment and migration in Tajikistan*.
48. Vasilyevich, R. S., & Khasanovich, R. A. . (2024). Emigration of Medical Personnel from Tajikistan Abroad: Causes and Consequences. *Migration Letters*, *21*(4), 1755–1767. https://doi.org/10.59670/ml.v21iS68489
49. Vyas, S., & Kumaranayake, L. (2006). Constructing socio-economic status indices: how to use principal components analysis. *Health Policy and Planning*, 21(6), 459-468. https://doi.org/10.1093/heapol/czl029
50. Wetzinger, J. (2024). Tajikistan's Higher Education System in Transformation. *International Higher Education*. https://doi.org/10.6017/895b9e0d.6f782d5a
51. World Bank. (2021). *Tajikistan overview: World Bank in Tajikistan*. https://www.worldbank.org/en/country/tajikistan/overview
52. *World Bank Migration Survey, 2013:* https://microdata.worldbank.org/index.php/catalog/2813/data-dictionary/F31/?offset=30
53. World Bank. (2023). *Economic outlook and labor market data for Tajikistan*. https://www.worldbank.org/en/country/tajikistan/overview
54. Wouterse, F. (2010). Migration from rural Burkina Faso: Effects on land and livestock ownership. *Journal of Development Economics,* *92*(2), 161-169. https://doi.org/10.1016/j.jdeveco.2009.06.002
55. Yang, D. (2011). Migrant remittances. *Journal of Economic Perspectives, 25*(3), 129–151. https://doi.org/10.1257/jep.25.3.129
56. Yang, D., & Choi, H. J. (2007). Are remittances insurance? Evidence from rainfall shocks in the Philippines. *World Bank Economic Review,* *21*(2), 219–248. https://doi.org/10.1093/wber/lhm003
57. Yormirzoev, M. (2017). *Factors influencing labor migration flows from Tajikistan to Russia*. Central Asian Journal of Migration. https://doi.org/10.14254/2071-789X.2017/10-3/5

Chapter-5: General Conclusion

In this dissertation we tried to explore the questions of aviation, regional economic development and migration in the context of case studies both in the USA and internationally.

Chapter-1 examines the dynamic causal relationship among the series of airport variables and regional economic development variables for the Denver MSA area and overall Colorado region - to estimate if there is a Granger causality between air transportation and regional development (and vice-versa). After applying various time series methods, the results reveal bi-directional, positive causality between *Total Domestic Flight Passengers* and *Total Business Entities* in Denver MSA, with larger effect of air transportation on regional development. Also, results demonstrate positive effect from *Domestic Flight Passengers* to *Total Employment* in Denver, but negative effect on *Total Employment* in Colorado, which suggests that domestic air transportation is important for local employment in the region, especially in a big hub city like Denver, but this effect fades out while spreading in more remote regions of Colorado. The policy implication from these results is that air transportation plays more important role in economic development than vice versa, and that effect is different between city-bub areas and more remote regions: in city-hub areas airports play as a strong catalysators of economic development, while in more remote areas of Colorado this effect is fading out. So, these results could be useful for policy-makers in forming their strategies of developing more remote/rural areas in Colorado. Yet, there is still a limitation with this result to be noted: Denver International Airport is a transportation hub that connects different cities not only in Colorado, so people might travel to/from Denver for transferring purposes to go to other places beside Colorado. As a result, transferring passengers might play a considerable but different role in local economic development. And due to data limitation transferring passengers were not considered in this estimation.

Chapter-2 explores the dynamics of inter-fuel substitution between biofuels and traditional fossil fuels in the general transportation sector and aviation sector in California. After using OLS and SUR models, results reveal that biofuels have not significantly changed traditional substitution and complementarity relationships in the general transportation sector, but they have become a viable alternative to kerosene-based jet fuel in aviation: increase in jet fuel price leads to considerable increase of SAF volumes consumed, demonstrating that SAF has become a valid substitute for a traditional jet fuel. So, this result demonstrates biofuels' growing potential to reduce dependency on fossil fuels in aviation. The limitation of this result is that due to data limitation not "pure" SAF prices, but their proxies, have been used in the analysis, which might be important factor in estimating the inter-substitutability between different fuels in the aviation sector.

Chapter-3 explores the migration phenomena in Tajikistan, analyzing what factors drive people to migrate from Tajikistan and to remit back to Tajikistan. By using Probit and Heckman models, results reveal that having networks and family arrangement of migration costs have a highest strong positive association with the probability of having migrants in the households, demonstrating that migration networks and family arranged costs of migration mostly are forming the migrations flows from Tajikistan to Russia. Another important result from this analysis reveals the characteristics of migration flows from Tajikistan: college education and language skills reveal a negative association with the probability of having migrants in the households, suggesting a Negative Selection migration process happening in Tajikistan, when less skilled workers tend to migrate to Russia from Tajikistan.

Appendix-A Chapter-2: Summary of Major Data Series

Figures A1-A7 provide a further summary of the data sources.

Raw data **Seasonally adjusted data**

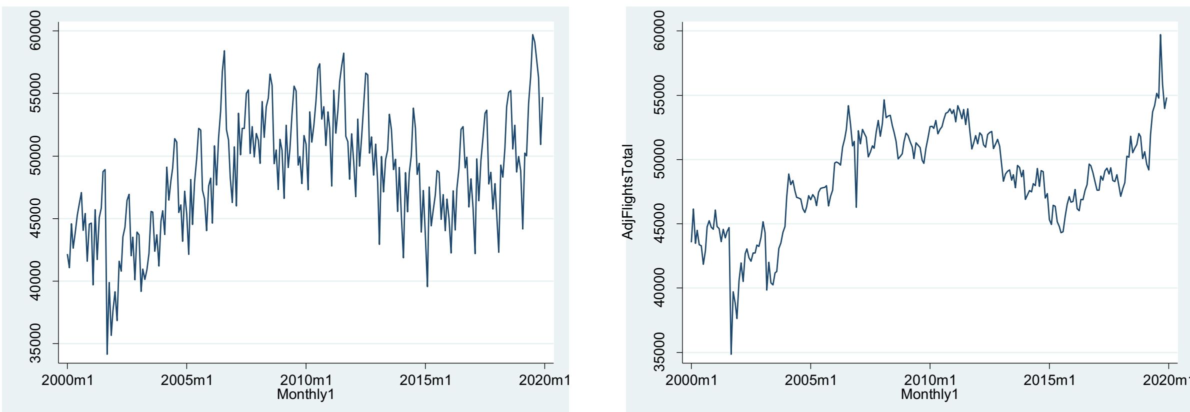


Figure-A1. Total Number of Flights, Denver International Airport
Source: Denver International Airport (DIA)

The total number of flights demonstrates an increasing trend over 2001-2010, with a slight drop in 2011-2015 following the 2008-2009 crisis.

Raw data **Seasonally adjusted data**

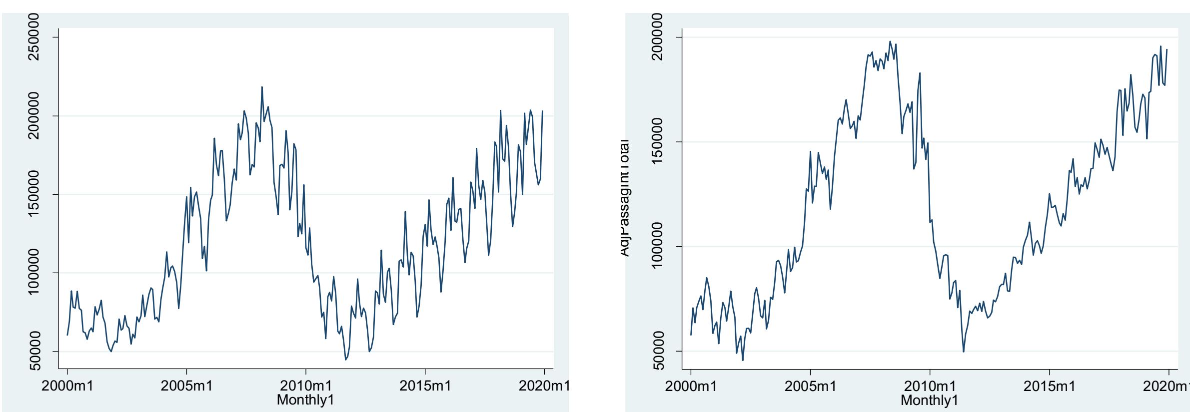


Figure-A2. Total Number of International Flight Passengers, Denver International Airport
Source: Denver International Airport (DIA)

The total number of international passengers demonstrates an increasing trend in 2002-2009 with a drop in 2010-2011 but a subsequent increasing trend in 2012-2019. The drop of international passengers in 2010-2011 may be associated with a crisis in Europe, though establishing this association is beyond the scope of this paper.

Raw data **Seasonally adjusted data**

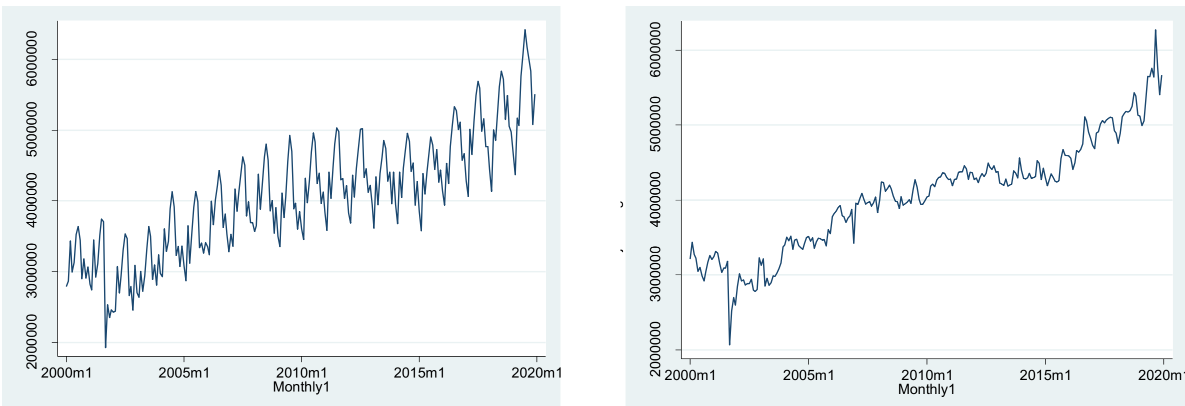


Figure-A3. Total Number of Domestic Flight Passengers, Denver International Airport
Source: Denver International Airport (DIA)

Domestic flight passengers demonstrate a stable increasing trend over time.

Raw data **Seasonally adjusted data**

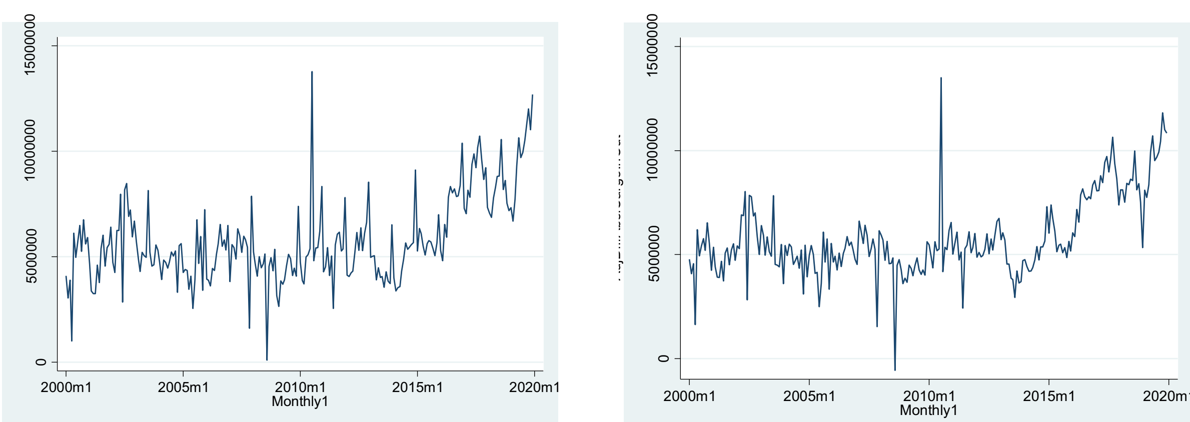


Figure-A4. Net Difference between In-Out Cargo (pounds), Denver International Airport
Source: Denver International Airport (DIA)

Net cargo demonstrates that the difference in inbound and outbound cargo over time is mainly stable in 2000-2014 though increasing since 2015. In 2010, there was much more inbound cargo brought to DIA than outbound cargo. These numbers are confirmed but precise reasons are unclear.

Raw data **Seasonally adjusted data**

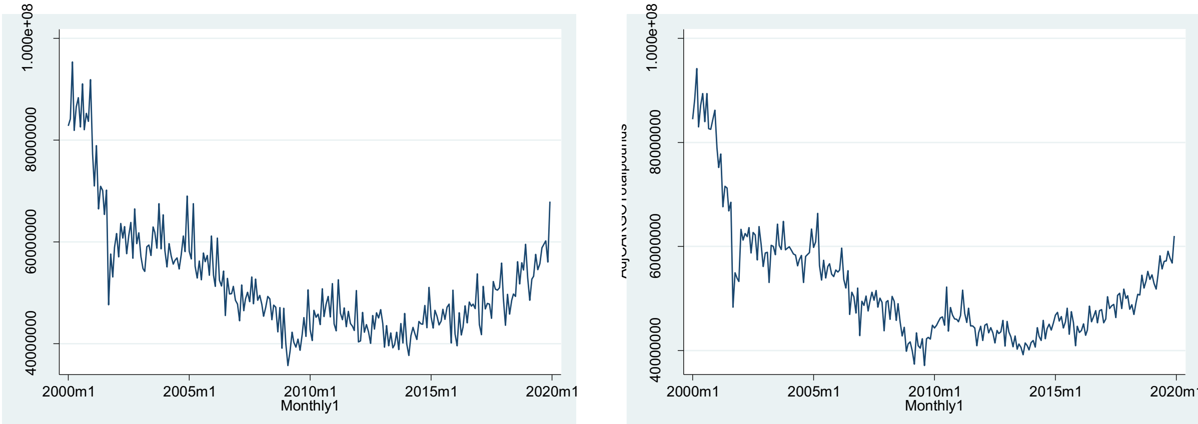


Figure-A5. Total Amount of Cargo (pounds), Denver International Airport
Source: Denver International Airport (DIA)

When we look at the total amount of cargo at DIA over the years, we see mainly a decreasing trend in 2000-2010, with a bit of increase in 2015-2019.

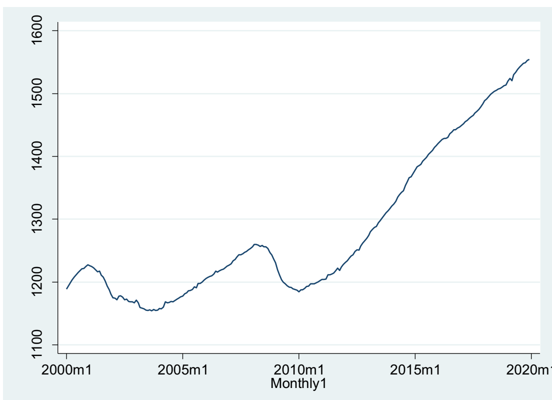
Raw data **Seasonally adjusted data**

Figure-A6. Number of Business Entities, Denver MSA

Source: State of Colorado

The number of businesses in Denver MSA demonstrates an increasing trend with the highest number of businesses emerging in 2019. This probably is associated with the overall enlarging of the business economy in the Denver MSA area.

Denver MSA
Raw data, already seasonally adjusted



Colorado Region
Raw data, already seasonally adjusted

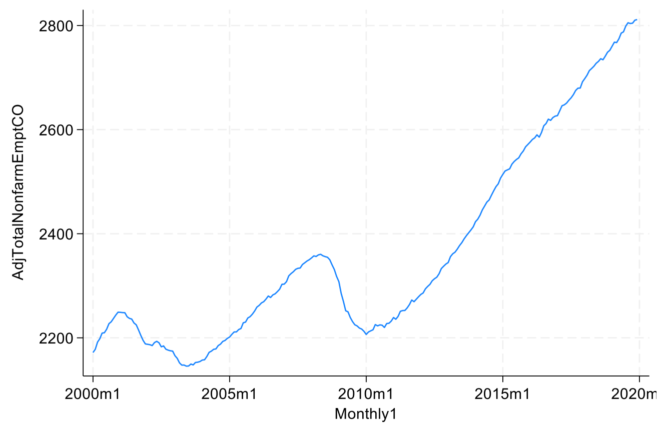


Figure-A7. Total Non-Farm Employment (thousands of people)
Source: US Bureau of Labor Statistics (BLS)

When we compare Total Non-Farm Employment in both Denver MSA and in Colorado overall, we see that both series look similar with highly increasing trends since 2011, with drops in 2004 and 2010.

Appendix-B Chapter-2: Diagnostic Test Results for Major Empirical Models

To estimate the validity of the estimated models several diagnostic tests are performed.

1. ARDL model results

Testing Pair-1: *Total Non-Farm Employment (Denver MSA and Colorado region) vs. Total Domestic Flights Passengers (DIA)*

Table-B1. Model Diagnostic Test Results

Denver MSA Region

Dependent Var.	Test	Lags	Statistics	P-value	Null Hypothesis	Result
Ln_Total Non-Farm Employment	Breusch-Godfrey LM test for autocorrelation	4	3.368	0.4982	H_0: no serial correlation	No autocorrelation
Ln_Total Non-Farm Employment	White Heteroskedasticity Test	4	74.00	0.0367	H_0: homoskedasticity	Heteroskedasticity detected
Ln_Total Domestic Flights Passengers	Breusch-Godfrey LM test for autocorrelation	4	4.731	0.3160	H_0: no serial correlation	No autocorrelation
Ln_Total Domestic Flights Passengers	White Heteroskedasticity Test	4	109.22	0.0000	H_0: homoskedasticity	Heteroskedasticity detected

Colorado Region

Dependent Var.	Test	Lags	Statistics	P-value	Null Hypothesis	Result
Ln_ Total Non-Farm Employment	Breusch-Godfrey LM test for autocorrelation	4	4.599	0.3310	H_0: no serial correlation	No autocorrelation Re-check
Ln_ Total Non-Farm Employment	White Heteroskedasticity Test	4	73.37	0.0036	H_0: homoskedasticity	Heteroskedasticity detected Re-check
Ln_Total Domestic Flights Passengers	Breusch-Godfrey LM test for autocorrelation	4	2.123	0.7132	H_0: no serial correlation	No autocorrelation
Ln_Total Domestic Flights Passengers	White Heteroskedasticity Test	4	6.27	0.9591	H_0: homoskedasticity	No heteroskedasticity

In the short-run models where *Total Non-Farm Employment* is defined as dependent variable, the regression for the underlying ARDL in both cases for Denver MSA and Colorado region fits well and the model is globally significant at 1% level. For the Denver MSA, the model also passes the diagnostic test against serial correlation (Durbin Watson test and Breusch-Godfrey test), but it does not pass the heteroscedasticity test (White Heteroskedasticity Test). For the Colorado region, the model passes the heteroscedasticity test (White Heteroskedasticity Test) but fails the diagnostic test against serial correlation (Durbin Watson test and Breusch-Godfrey test.

In the reverse short-run model, where *Total Domestic Flights Passengers* is defined as dependent variable, the regression for the underlying ARDL in both cases for Denver MSA and Colorado region fits well and the model is globally significant at 1% level. For Denver, the model also passes the diagnostic test against serial correlation but does not pass the heteroscedasticity test. For the Colorado region, the model passes the diagnostic test against serial correlation (Durbin Watson test and Breusch-Godfrey test), also it passes the heteroscedasticity test (White Heteroskedasticity Test).

Testing Pair-2: *Total Non-Farm Employment (Denver MSA and Colorado region) vs. Difference of Inbound-Outbound Cargo (DIA)*

Table-B2. Model Diagnostic Test Results

Denver MSA Region

Dependent Var.	Test	Lags	Statistics	P-value	Null Hypothesis	Result
Ln_Total Non-Farm Employment in Denver MSA	Breusch-Godfrey LM test for autocorrelation	4	1.363	0.8506	H_0: no serial correlation	No autocorrelation
Ln_Total Non-Farm Employment in Denver MSA	White Heteroskedasticity Test	4	60.73	0.0478	H_0: homoskedasticity	Heteroskedasticity detected
Ln_Net In-Out Cargo (DIA)	Breusch-Godfrey LM test for autocorrelation	4	5.710	0.1266	H_0: no serial correlation	No autocorrelation

Ln_Net In-Out Cargo (DIA)	White Heteroskedasticity Test	4	22.31	1.0000	H_0: homoskedasticity	No heteroskedasticity

Colorado Region

Dependent Var.	Test	Lags	Statistics	P-value	Null Hypothesis	Result
Ln_Total Non-Farm Employment in Colorado region	Breusch-Godfrey LM test for autocorrelation	4	29.548	0.0000	H_0: no serial correlation	Autocorrelation detected
Ln_Total Non-Farm Employment in Colorado region	White Heteroskedasticity Test	4	57.53	0.0829	H_0: homoskedasticity	No heteroskedasticity
Ln_Net In-Out Cargo (DIA)	Breusch-Godfrey LM test for autocorrelation	4	3.267	0.5142	H_0: no serial correlation	No autocorrelation
Ln_Net In-Out Cargo (DIA)	White Heteroskedasticity Test	4	8.78	0.9853	H_0: homoskedasticity	No heteroskedasticity

In the short-run models where *Total Non-Farm Employment* is defined as dependent variable, the regressions for the underlying ARDL for both the Denver MSA and Colorado region overall fit well and the models are globally significant at 1% level. For Denver MSA region, the model also passes the diagnostic test against serial correlation (Durbin Watson test and Breusch-Godfrey test), but it does not pass the heteroscedasticity test (White Heteroskedasticity Test). For the Colorado region, the model passes the heteroscedasticity test (White Heteroskedasticity Test) but fails the diagnostic test against serial correlation (Durbin Watson test and Breusch-Godfrey test.

In the reverse short-run model, where *Difference of Inbound-Outbound Cargo* is defined as dependent variable, the regression for the underlying ARDL in both cases for Denver MSA and Colorado region fits well and the model is globally significant at 1% level. The models also pass the diagnostic test against serial correlation (Durbin Watson test and Breusch-Godfrey test) and the heteroscedasticity test White Heteroskedasticity Test).

Testing Pair-3: ***Total International Flights Passengers (DIA) vs. Total Registered Businesses in Denver MSA***

Table-B3. Model Diagnostic Test Results

Dependent Var.	Test	Lags	Statistics	P-value	Null Hypothesis	Result
Ln_ Total Registered Businesses in Denver MSA	Breusch-Godfrey LM test for autocorrelation	4	7.722	0.1029	H_0: no serial correlation	No autocorrelation
Ln_ Total Registered Businesses in Denver MSA	White Heteroskedasticity Test	4	26.47	0.1509	H_0: homoskedasticity	No Heteroskedasticity
Ln_Total International Flights Passengers	Breusch-Godfrey LM test for autocorrelation	4	3.186	0.5272	H_0: no serial correlation	No autocorrelation
Ln_Total International Flights Passengers	White Heteroskedasticity Test	4	45.96	0.1017	H_0: homoskedasticity	No Heteroskedasticity

In the first short-run model, the regression for the underlying ARDL fits very well and the model is globally significant at 1% level. It also passes the diagnostic test against serial correlation (Durbin Watson test and Breusch-Godfrey test) and the heteroscedasticity test (White Heteroskedasticity Test). For reverse short-run model, the regression for the underlying ARDL fits very well and the model is globally significant at 1% level. It also passes the diagnostic test against serial correlation (Durbin Watson test and Breusch-Godfrey test) and the heteroscedasticity test (White Heteroskedasticity Test).

2. Causality Test Results through VECM approach

Testing Pair-4: ***Total Registered Businesses in Denver MSA vs. Total Domestic Flights Passengers (DIA)***

Table-B4. Model Diagnostic Test Results

Dependent Var.	Test	Lags	Statistics	P-value	Null Hypothesis	Result
Ln_ Total Registered Businesses in Denver MSA	Lagrange Multiplier for autocorrelation	4	4.1091	0.3914	H_0: no serial correlation	No autocorrelation
Ln_ Total Domestic Flights Passengers in DIA	Lagrange Multiplier for autocorrelation	4	4.1091	0.3914	H_0: no serial correlation	No autocorrelation

In both models, the regressions for the underlying VECM fit very well and the models are globally significant at 1% level. They also pass the diagnostic test against serial correlation (Lagrange Multiplier test).

Appendix-C Chapter-3: Correlation Matrices

Table-C1. Correlation matrix of volumes and prices of primary fuels in the transportation sector of California, 1970-2021

	Aviation Gasoline Volume	Distillate Fuel Oil Volume	HGL (Propane) Volume	Jet Fuel Volume	Motor Gasoline Volume	Residual Fuel Oil Volume	Electricity Volume	Aviation Gasoline Price	Distillate Fuel Oil Price	HGL (Propane) Price	Jet Fuel Price	Motor Gasoline Price	Residual Fuel Oil Price	Electricity Price
Aviation Gasoline Volume	1													
Distillate Fuel Oil Volume	-0.83	1												
HGL (Propane) Volume	0.09	-0.24	1											
Jet Fuel Volume	-0.56	0.73	-0.02	1										
Motor Gasoline Volume	-0.74	0.89	-0.10	0.88	1									
Residual Fuel Oil Volume	-0.03	-0.26	0.65	-0.17	-0.18	1								
Electricity Volume	-0.80	0.93	-0.28	0.67	0.89	-0.35	1							
Aviation Gasoline Price	-0.65	0.54	0.01	0.01	0.35	0.15	0.57	1						
Distillate Fuel Oil Price	-0.75	0.70	-0.18	0.20	0.51	-0.02	0.73	0.94	1					
HGL (Propane) Price	-0.70	0.71	0.14	0.51	0.74	0.14	0.72	0.65	0.70	1				
Jet Fuel Price	-0.59	0.43	0.11	0.00	0.33	0.26	0.51	0.96	0.89	0.68	1			
Motor Gasoline Price	-0.59	0.45	-0.15	-0.08	0.27	0.08	0.54	0.94	0.92	0.52	0.92	1		
Residual Fuel Oil Price	-0.53	0.41	-0.12	-0.12	0.24	0.00	0.50	0.93	0.88	0.54	0.90	0.91	1	
Electricity Price	-0.19	0.03	0.43	-0.32	-0.10	0.53	-0.01	0.55	0.36	0.29	0.52	0.46	0.44	1

Source: Authors calculations based on data from the U.S. Energy Information Administration (EIA).

Table-C2. Correlation matrix of volumes and prices of primary aviation fuels in the aviation sector of California, 2013-2019

	Jet Fuel Volume	Jet Fuel Price	SAF Volume	SAF Price
Jet Fuel Volume	1			
Jet Fuel Price	0.1504	1		
SAF Volume	-0.3374	0.4967	1	
SAF Price	-0.1767	-0.126	0.1817	1

Source: Authors calculations based on data from the U.S. Energy Information Administration (EIA) and the U.S. Environmental Protection Agency (EPA).

Appendix-D1 Chapter-4: Correlation Matrix

Table-D1. Correlation Matrix

	D. If HH have Migrant (Y=1, N=0)	*HH Size*	*Proportion of Children (%)*	*Proportion of Seniors (%)*	*D. If HH has a Member with College Education (Y=1, N=0)*	*Proportion of HH Members with Intermediate Russian Skills (%)*	*Annual Income of HH (Tajik Somoni)*	*D. If Networks Arranged (Y=1, N=0)*	*D. If HH has Arranged Costs of Migration (Y=1, N=0)*
D. If HH has Migrant (Y=1, N=0)	1								
HH Size	0.229	1							
Proportion of Children (%)	-0.024	0.287	1						
Proportion of Seniors (%)	-0.078	-0.166	-0.346	1					
D. If HH has a Member with College Education (Y=1, N=0)	-0.144	-0.045	-0.090	0.042	1				
Proportion of HH Members with Intermediate Russian Skills (%)	-0.131	-0.345	-0.415	0.156	0.346	1			
Annual Income of HH (Tajik Somoni)	0.022	0.050	0.000	-0.018	0.026	0.005	1		
D. If Networks Arranged (Y=1, N=0)	0.389	0.248	0.003	-0.123	-0.120	-0.108	0.024	1	
D. If HH has Arranged Costs of Migration (Y=1, N=0)	0.302	0.129	0.023	-0.065	-0.102	-0.097	-0.026	0.268	1

Note: No major correlations are detected for main variables.

Source: Authors calculations.

Appendix-D2 Chapter-4: Probit Model, using Wealth Index

Table-D2. Probit Model Output, using Wealth Index

	(Probit: All)	(Probit: Urban)	(Probit: Rural)
VARIABLES	**Prob. of HH to have Migrant (Y=1, N=0)**	**Prob. of HH to have Migrant (Y=1, N=0)**	**Prob. of HH to have Migrant (Y=1, N=0)**
HH Size	0.0705***	0.0831***	0.0524***
	(0.00998)	(0.0154)	(0.0134)
Proportion of Children (%)	-0.00818***	-0.00897***	-0.00619***
	(0.00148)	(0.00208)	(0.00219)
Proportion of Seniors (%)	-0.00419**	-0.00531**	-0.00263
	(0.00196)	(0.00257)	(0.00346)
1D. If HH has a Member with College Education (Y=1, N=0)	-0.258***	-0.186**	-0.160
	(0.0600)	(0.0797)	(0.0999)
Proportion of HH Members with Intermediate Russian Skills (%)	-0.00246**	-0.000342	-0.000238
	(0.000963)	(0.00146)	(0.00147)
Annual Income of HH (without Remittances) (Tajik Somoni)	2.41e-06	-5.00e-07	3.23e-06
	(3.20e-06)	(4.03e-06)	(4.56e-06)
2. Wealth Index (below medium)	0.0278	-0.104	0.132
	(0.0905)	(0.103)	(0.280)
3. Wealth Index (medium)	-0.0716	-0.369***	-0.124
	(0.0903)	(0.118)	(0.268)
4. Wealth Index (upper medium)	0.0448	-0.0382	-0.121
	(0.0949)	(0.125)	(0.271)
5. Wealth Index (rich)	0.0590	-0.204	-0.0695
	(0.0934)	(0.130)	(0.268)
1D. If Networks Arranged (Y=1, N=0)	0.825***	0.910***	0.691***
	(0.0570)	(0.0818)	(0.0814)
1D. If HH has Arranged Costs of Migration (Y=1, N=0)	0.959***	1.154***	0.815***
	(0.101)	(0.157)	(0.133)
Constant	-0.890***	-1.166***	-0.612**
	(0.128)	(0.178)	(0.295)
Observations	2,970	1,814	1,156

*Note: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.*

Source: Authors calculations.

Appendix-D3 Chapter-4: Probit Model, using Consumption Quantiles

Table-D3. Robustness Check Probit Model Output, using Consumption Quantiles

	(Probit: All)	(Probit: Urban)	(Probit: Rural)
VARIABLES	**Prob. of HH to have Migrant (Y=1, N=0)**	**Prob. of HH to have Migrant (Y=1, N=0)**	**Prob. of HH to have Migrant (Y=1, N=0)**
HH Size	0.0651***	0.0613***	0.0452***
	(0.00983)	(0.0152)	(0.0134)
Proportion of Children (%)	-0.00849***	-0.00921***	-0.00629***
	(0.00149)	(0.00209)	(0.00221)
Proportion of Seniors (%)	-0.00439**	-0.00609**	-0.00273
	(0.00196)	(0.00261)	(0.00346)
1D. If HH has a Member with College Education (Y=1, N=0)	-0.240***	-0.123	-0.163
	(0.0602)	(0.0810)	(0.0998)
Proportion of HH Members with Intermediate Russian Skills (%)	-0.00236**	-6.61e-05	0.000271
	(0.000953)	(0.00146)	(0.00146)
Annual Income of HH (without Remittances) (Tajik Somoni)	2.32e-06	1.37e-07	3.51e-06
	(3.26e-06)	(4.09e-06)	(4.64e-06)
Consumption Quantiles			
2. Quintile_2013_TJ	-0.0258	-0.0424	-0.0849
	(0.0893)	(0.123)	(0.134)
3. Quintile_2013_TJ	-0.142	-0.338**	-0.107
	(0.0894)	(0.131)	(0.130)
4. Quintile_2013_TJ	-0.0391	-0.209*	-0.0493
	(0.0873)	(0.125)	(0.129)
5. Quintile_2013_TJ	-0.181**	-0.285**	-0.212
	(0.0861)	(0.119)	(0.131)
1D. If Networks Arranged (Y=1, N=0)	0.839***	0.906***	0.695***
	(0.0567)	(0.0812)	(0.0811)
1D. If HH has Arranged Costs of Migration (Y=1, N=0)	0.948***	1.093***	0.803***
	(0.101)	(0.156)	(0.133)
Constant	-0.767***	-1.023***	-0.554***
	(0.130)	(0.190)	(0.187)
Observations	2,970	1,814	1,156

*Note: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.*

Source: Authors calculations.

Appendix-D4 Chapter-4: Wealth Index Construction

D4. Determination of Household Wealth Index

The households were from both urban and rural areas and from different income categories. Therefore, the determination of their household wealth index should take this difference into account as the variables (assets) used to measure wealth may vary for urban and rural dwellers. The composite index is constructed using Principal Component Analysis, which gives the scores based on ownership of different durable assets by households: number of rooms occupied by household, availability of individual heating system, water supply system and its status (operational or not), sewage system, status of power system (operational or not), availability of electric cooking appliances. In the context of Tajikistan, these variables reflect the wealth level of households, since after 1990 power supply, water supply, heating systems are outdated and had not been renovated systematically by the local governments because of lack of budget, and population try to renovate them privately (World Bank Report: Drinking Water, Sanitation and Hygiene Conditions in Tajikistan, 2017).

According to Filmer and Pritchett (2001), and as further supported by Minujin and Hee Bang (2002), McKenzie (2005), Vyas and Kumaranayake (2006), and Labonne, Biller, and Chase (2007), Principal Component Analysis (PCA) is commonly used in economics and public policy research to estimate wealth effects. This approach estimates relative wealth by using the first principal component, which is expressed as a linear combination for each household *i*:

$$y_i = \alpha_1 \left(\frac{x_1 - \bar{x}_1}{s_1}\right) + \alpha_2 \left(\frac{x_2 - \bar{x}_2}{s_2}\right) + \cdots + \alpha_k \left(\frac{x_k - \bar{x}_k}{s_k}\right) \qquad \text{(A1)}$$

Here, $\bar{x}_k$ and s_k represent the mean and standard deviation of each asset x_k, while α represents the weight of each variable x_k in the first principal component. By construction, this principal component has a mean of zero and a variance equal to the largest eigenvalue λ of the correlation matrix for x. This component, or wealth index, y, prioritizes assets that show significant variation across households, assigning a weight of zero to assets universally owned (McKenzie 2005). Consequently, the wealth index can produce both positive and negative values.

Following Vyas and Kumaranayake (2006), all variables were initially converted into binary indicators (1 = Yes, 0 = No) to signify the presence of household assets. Separate weights (factor scores) were calculated for urban and rural areas within each country. A "relative wealth" variable was then created in the combined dataset to account for the differing distributions of assets in urban and rural areas, thus better reflecting the economic conditions within each country.

Source:
1. Determination of household wealth index, BMJ Open.
2. Cordova, A. (2008). Methodological Note: Measuring Relative Wealth using Household Asset Indicators, AmericasBarometer Insights: 2008 (No.6). https://www.vanderbilt.edu/lapop/insights/I0806en_v2.pdf