Essays on the Search and Matching of Economic Agents in Labor and Transportation Markets

Yiliang Li

University College
University of Oxford

A Thesis Submitted for the Degree Of
Doctor of Philosophy

Trinity 2022

Abstracts

The Agricultural Exodus in the Philippines: Are Wage Differentials Driving the Process?

Lagging labor reallocations outside agriculture amid sustained low agricultural productivity have been a key feature in the Philippines over the past 15 years. An analysis of the labor adjustments in and out of agriculture shows that a variety of factors have influenced this process. We find that the widening of wage differentials with non-agricultural sectors, improvements in labor market efficiency, and better transport infrastructure are largely associated with growing outflows of labor from agriculture, whilst the lack of post-primary education and the presence of agricultural clusters hinder such outflows. In contrast to the traditional view that agricultural employment outflows are largely driven by productivity differences and wage differentials, our results emphasize the roles of education and transport infrastructure in facilitating labor reallocations from agriculture to non-agriculture.

JEL Codes: E24, J24, J43, J64

Keywords: Agricultural Employment, Wage Differentials, Labor Market Efficiency, Transport Infrastructure, Education, Search and Matching

A Big Data Approach to Cargo Type Prediction and Its Implications for Oil Trade Estimation

To estimate global crude oil trade flows, current research either considers only crude oil tankers, or simply applies external information to distinguish between crude and refined product oil cargoes transported by coated product tankers; these

limitations often reduce an estimation's accuracy or compromise replicability. Our methodology directly addresses these issues by applying the random forest (RF) ensemble learning technique to Automatic Identification System (AIS) data in order to predict the cargo types of coated product tankers. By leveraging domain knowledge, we construct a set of unique input variables for the RF model, and use its predictions to quantify the global crude oil trade in a more accurate manner. Our estimation shows that coated product tankers were responsible for approximately 8% of global seaborne crude oil trade from 2017-2020. Further, unanticipated variations in the crude oil volume carried by these tankers are consistent with several major historical oil trade disruptions. Our study further extends current applications of AIS data in the domains of operations management and maritime transportation, and facilitates the exploration of the more minute characteristics of oil transportation. The resulting shipping dataset and associated decomposition strategy also enable relevant stakeholders to quickly identify emerging trade flow risks and adapt more effectively.

JEL Codes: E32, F10, Q40, R40

Keywords: Cargo Type Prediction, Oil Trade Estimation, Automatic Identification System, Random Forest

The Congestion Effect of Oil Transportation and Its Trade Implications

This paper uses a unique dataset, encompassing the entire product tanker fleet from 2017-2020, to study congestion in oil transportation markets. From this we establish three novel facts: (1) congestion significantly reduces tankers' contracting probability; (2) tankers switch frequently between markets; and, (3) oil trade shocks propagate in the economy through a transportation network. We develop a novel search and matching model by incorporating these facts with the interactions between oil exporters, tankers, and a final good producer. We show that shipping market congestion attenuates over one third of the volatility in oil trade and final good production, while increasing price fluctuations by more than tenfold. We also show that the endogenous adjustment of the oil transportation network and the resulting congestion effect are almost irrelevant for aggregate fluctuations; it is the exogenous oil trade shocks that generate large fluctuations.

JEL Codes: E32, F10, Q43, R41

Keywords: Oil Transportation, Oil Trade and Price, Congestion, Transportation Network, AIS, Search and Matching

Essays on the Search and Matching of Economic Agents in Labor and Transportation Markets



Yiliang Li

University College

University of Oxford

A Thesis Submitted for the Degree Of

Doctor of Philosophy

Trinity 2022

Declaration

Supervisor: Prof. Francesco Zanetti.

Word Count: Approximately 44,000 words.

Authorship: The work in this thesis is based on my research as a DPhil candidate at the University of Oxford between October 2018 and June 2022. No part of this thesis has been submitted for a degree at the University of Oxford or any other university. The chapter, "The Agricultural Exodus in the Philippines: Are Wage Differentials Driving the Process?", represents my own work during my internship at the International Monetary Fund (IMF) under the supervision of Dr. Eugenio Cerutti. The paper is now circulated as an IMF Working Paper (No. 2021/220). The chapter, "A Big Data Approach to Cargo Type Prediction and Its Implications for Oil Trade Estimation", is completed in collaboration with Prof. Xiwen Bai, Qi Wang, and Zhongjun Ma, all of whom are affiliated with Tsinghua University. I am the lead of the project and responsible for all the written work. The chapter, "The Congestion Effect of Oil Transportation and Its Trade Implications", is my job market paper, coauthored with Prof. Xiwen Bai. I am the lead of the project and responsible for all the written, analytical, and numerical work except for the initial data cleaning and processing.

Funding: I acknowledge the financial support from the Two-Year Department of Economics Doctoral Bursary for the academic years 2018-19 and 2019-20, and that from the Final-Year Departmental Doctoral Bursary for the academic year 2020-21.

Declaimer: The opinions expressed in "The Agricultural Exodus in the Philippines: Are Wage Differentials Driving the Process?" are solely the responsibility of the authors and should not be interpreted as reflecting those of the IMF, its Executive Board, or IMF management.

Abstracts

The Agricultural Exodus in the Philippines: Are Wage Differentials Driving the Process?

Lagging labor reallocations outside agriculture amid sustained low agricultural productivity have been a key feature in the Philippines over the past 15 years. An analysis of the labor adjustments in and out of agriculture shows that a variety of factors have influenced this process. We find that the widening of wage differentials with non-agricultural sectors, improvements in labor market efficiency, and better transport infrastructure are largely associated with growing outflows of labor from agriculture, whilst the lack of post-primary education and the presence of agricultural clusters hinder such outflows. In contrast to the traditional view that agricultural employment outflows are largely driven by productivity differences and wage differentials, our results emphasize the roles of education and transport infrastructure in facilitating labor reallocations from agriculture to non-agriculture.

JEL Codes: E24, J24, J43, J64

Keywords: Agricultural Employment, Wage Differentials, Labor Market Efficiency, Transport Infrastructure, Education, Search and Matching

A Big Data Approach to Cargo Type Prediction and Its Implications for Oil Trade Estimation

To estimate global crude oil trade flows, current research either considers only crude oil tankers, or simply applies external information to distinguish between crude and refined product oil cargoes transported by coated product tankers; these limitations often reduce an estimation's accuracy or compromise replicability. Our methodology directly addresses these issues by applying the random forest (RF) ensemble learning technique to Automatic Identification System (AIS) data in order to predict the cargo types of coated product tankers. By leveraging domain knowledge, we construct a set of unique input variables for the RF model, and use its predictions to quantify the global crude oil trade in a more accurate manner. Our estimation shows that coated product tankers were responsible for approximately

8% of global seaborne crude oil trade from 2017-2020. Further, unanticipated variations in the crude oil volume carried by these tankers are consistent with several major historical oil trade disruptions. Our study further extends current applications of AIS data in the domains of operations management and maritime transportation, and facilitates the exploration of the more minute characteristics of oil transportation. The resulting shipping dataset and associated decomposition strategy also enable relevant stakeholders to quickly identify emerging trade flow risks and adapt more effectively.

JEL Codes: E32, F10, Q40, R40

Keywords: Cargo Type Prediction, Oil Trade Estimation, Automatic Identification System, Random Forest

The Congestion Effect of Oil Transportation and Its Trade Implications

This paper uses a unique dataset, encompassing the entire product tanker fleet from 2017-2020, to study congestion in oil transportation markets. From this we establish three novel facts: (1) congestion significantly reduces tankers' contracting probability; (2) tankers switch frequently between markets; and, (3) oil trade shocks propagate in the economy through a transportation network. We develop a novel search and matching model by incorporating these facts with the interactions between oil exporters, tankers, and a final good producer. We show that shipping market congestion attenuates over one third of the volatility in oil trade and final good production, while increasing price fluctuations by more than tenfold. We also show that the endogenous adjustment of the oil transportation network and the resulting congestion effect are almost irrelevant for aggregate fluctuations; it is the exogenous oil trade shocks that generate large fluctuations.

JEL Codes: E32, F10, Q43, R41

Keywords: Oil Transportation, Oil Trade and Price, Congestion, Transportation Network, AIS, Search and Matching

Acknowledgements

I am deeply indebted to my supervisor, Francesco Zanetti, for his invaluable guidance throughout my doctoral research. This thesis would never be possible had I not received his indispensable advice along the way. I am always grateful for his support and inspiration every time I am bogged down in the labyrinth of life; as Albert Schweitzer nicely puts it, at times when our own light goes out, it is always rekindled by a spark from another person.

I would like to thank Xiwen Bai, Eugenio Cerutti, and Petr Sedláček, who have given me extremely helpful advice on the specific chapters of this thesis. I could not have undertaken this endeavor without their continuous help, patience, and most importantly, great confidence in me. In addition, my work has benefited from valuable discussions with the faculty members at Oxford, my supervisors and colleagues at both the World Bank and International Monetary Fund, who generously provided knowledge and expertise.

I am also thankful to my friends who have always been there for me, during both the uphill and downhill patches of my life, whether nearby or faraway. Special thanks go to Xu Dong, Shihang Hou, Xiyu Jiao, Zhuhan Jin, Wu Le, Boyang Liu, Rajssa Mechelli, Sweta Saxena, Tatjana Schulze, Chenchuan Shi, Haikun Zhan, Ning Zhang, Wenzhang Zhang, Yifan Zhang, and Yuwei Zheng for being a generous source of joy, encouragement, and companionship. I would also like to extend my sincere thanks to University College for providing me with a place I could call home, and to Wadham College for giving me an opportunity to realize my enthusiasm for teaching during the past four years.

Lastly, I would be remiss in not mentioning my parents. No words can describe my heartfelt gratitude towards them for their enduring love and support, and their belief in me has kept my spirits and motivation high in every part of this journey.

Contents

List of Figures **xiii**

List of Tables **xv**

1 Introduction **1**

2 The Agricultural Exodus in the Philippines: Are Wage Differentials Driving the Process? **9**

2.1 Introduction 9
2.2 Labor Market in the Philippines 14
2.2.1 Data 14
2.2.2 Time Series Trends of the EN and NE Rates 16
2.2.3 'Push' and 'Pull' Drivers of Intersectoral Labor Reallocations 19
2.3 Empirical Analysis 30
2.3.1 Linear Probability Model 30
2.3.2 Baseline Regression Results 31
2.3.3 Robustness Checks 34
2.4 Which Drivers Have the Largest Impact? 41
2.5 Conclusion 43

3 A Big Data Approach to Cargo Type Prediction and Its Implications for Oil Trade Estimation **45**

3.1 Introduction 45
3.2 Literature Review 50
3.3 Data 53
3.4 Identification of Loaded Trips 54
3.4.1 Trips 54
3.4.2 Loaded and Unloaded Trips 54
3.5 Estimation of Cargo Types 56
3.5.1 Applying Existing Fixtures 57
3.5.2 Applying the Random Forest Prediction Model 57
3.6 Evaluation of Prediction Results 68
3.6.1 CPP vs. DPP Trips 68

3.6.2 Crude vs. Dirty Refined Product Oil Trips 72
3.7 Application of Prediction Results 73
3.8 Conclusion 78

4 The Congestion Effect of Oil Transportation and Its Trade Implications 79
4.1 Introduction 79
4.2 Oil Transportation 85
4.2.1 Overview 85
4.2.2 Data 86
4.2.3 Novel Facts 89
4.3 Baseline Model 98
4.3.1 Structure 98
4.3.2 Equilibrium 110
4.3.3 Parameterization 111
4.4 Model Implications 117
4.4.1 Impulse Response Functions 118
4.4.2 Counterfactual Exercises 121
4.5 Concluding Remarks 130

5 Conclusion 133

Appendices

A Appendices to "The Agricultural Exodus in the Philippines: Are Wage Differentials Driving the Process?" 139
A.1 Industry Classification 139
A.2 Theoretical Model 140
A.2.1 Environment 140
A.2.2 Value Functions 141
A.2.3 Surplus Sharing and Separation Decision 143
A.2.4 Partial Equilibrium in the Agricultural Labor Market 144
A.2.5 Comparative Statics 148
A.3 End Behaviors of the EN and NE Rates 152
A.4 Additional Tables and Figures 155

B Appendices to "A Big Data Approach to Cargo Type Prediction and Its Implications for Oil Trade Estimation" 163
B.1 Additional Tables and Figures 163
B.2 Alternative Prediction Models 165
B.2.1 Logistic Regression 165
B.2.2 SVM 167
B.2.3 Neural Network 167

C Appendices to "The Congestion Effect of Oil Transportation and Its Trade Implications" 169
C.1 The Presence of Search Frictions 169
C.2 Construction of the Weather Instruments 170
C.3 VAR Evidence of Shipping Market Congestion 170
C.4 Derivation of the Integrated Value Functions 173
C.5 Derivation of Commissions 174
C.6 Steady-State Existence 176
C.7 Converting Model Moments to Data Moments 178
C.8 MCMC Approach to SMM Estimation 179
C.9 Recovering Data-Consistent Trade Shocks 181
C.10 Baseline vs. Counterfactual Exercises 182
C.11 Counterfactual: Idling of Ships 182
C.12 Counterfactual: Free-Entry of Potential Exporters 187
C.13 Natural Experiments 189
C.13.1 IMO 2020 189
C.13.2 2021 Suez Canal Obstruction 191
C.13.3 An Unsanctioned Iran 192
C.14 Additional Impulse Response Functions 193

References 197

List of Figures

2.1 Agricultural performance in selected ASEAN countries 10
2.2 Labor market composition 17
2.3 Time series trends of the labor market transition rates 18
2.4 Job separation and creation in the agricultural labor market 20
2.5 The influence of higher non-agricultural productivity on job separation and finding 22
2.6 The influence of lower flow outside value for agricultural workers on job separation and finding 24
2.7 Real wage differentials 25
2.8 Labor market efficiency index 27
2.9 National road density 28
2.10 Share of employed workers with at least high school education . . . 29
2.11 Time series trends of the EE rates 37
2.12 Economic significance 42

3.1 Draught ratio and state classification 56
3.2 The likelihood of a CPP trip and trip distance 60
3.3 Geographic dispersion of oil trade 61
3.4 Freight earnings spread 63
3.5 Proportions of CPP (DPP) trips completed by the top 20 shipowners 64
3.6 Basic structure of an RF prediction Model 67
3.7 MDA values of the input variables (1st stage) 70
3.8 Global crude oil trade volume carried by LR tankers from 2017W5-2020W48 75
3.9 Remainder component of STL with $n_s = periodic$ 76

4.1 Geographic concentration of oil trade 92
4.2 Fluctuations in tanker utilization and major oil trade events from 2017W5-2020W48 95
4.3 IRFs to a positive crude oil trade shock 119
4.4 IRFs to a positive crude oil trade shock: Baseline vs. no ballasting . 127

A.1 The influence of lower agricultural productivity on job separation and finding . . . 157
A.2 The influence of better matching in agriculture on job separation and finding . . . 157
A.3 The influence of weaker bargaining of agricultural workers on job separation and finding . . . 158
A.4 AMEs of high school education, by region . . . 158
A.5 AMEs of high school education, by age . . . 159
A.6 AMEs of high school education, by gender . . . 159
A.7 AMEs of college education, by region . . . 160
A.8 AMEs of college education, by age . . . 160
A.9 AMEs of college education, by gender . . . 161
A.10 AMEs of road density, by region . . . 161

B.1 MDA values of the input variables (2nd stage) . . . 164
B.2 Remainder components of STL with different values of n_s . . . 165
B.3 Basic structure of a neural network . . . 168

C.1 Under-utilization of empty ships . . . 170
C.2 Empirical IRFs of average contracting probability to a tanker under-utilization shock . . . 172
C.3 IRFs to a positive crude oil trade shock: Baseline vs. idling . . . 186
C.4 IRFs to a positive crude oil trade shock: Baseline vs. free-entry . . . 188
C.5 Freight earnings amid the Suez Canal blockage . . . 192
C.6 IRFs to a positive product oil trade shock . . . 194
C.7 IRFs to a positive product oil trade shock: baseline vs. no ballasting 194
C.8 IRFs to a positive product oil trade shock: Baseline vs. idling . . . 195
C.9 IRFs to a positive product oil trade shock: Baseline vs. free-entry . 195

List of Tables

2.1 AMEs in the baseline LPM 32
2.2 AMEs in the LPM with cross-region real wage differentials 36
2.3 AMEs in the LPM for EE transitions 38
2.4 AMEs in the LPM for transitions from agriculture to other low-skilled industries 40

3.1 Summary statistics of input variables (1st stage) 59
3.2 Hyper-parameters in the RF model 68
3.3 Confusion matrix for the test set (1st stage) 69
3.4 Summary of the prediction results on the test set (1st stage) 72
3.5 Confusion matrix for the test set (2nd stage) 73

4.1 Estimation of the aggregate matching function 93
4.2 Summary statistics of the market switching probability 94
4.3 Model parameters and assigned values 112
4.4 Data vs. model moments from the SMM estimation 118
4.5 Baseline vs. counterfactual exercises 123

A.1 AMEs in the LPM with time dummy representing 2005Q2-2008Q1 . 153
A.2 AMEs in the LPM with time dummy representing 2017Q1-2019Q4 . 154
A.3 AMEs in the LPM with clustered standard errors at the regional level 155
A.4 AMEs of the proposed 'push' and 'pull' factors in a logit model . . 156

B.1 Summary statistics of input variables (2nd stage) 163
B.2 Summary of the prediction results on the test set (2nd stage) 164
B.3 Parameter estimates of the logistic regression model 166

C.1 Extended comparison: Baseline vs. counterfactual exercises 183
C.2 Baseline vs. natural experiments 190

1 Introduction

This thesis comprises three chapters which discuss the search and matching behaviors of economic agents in the labor and oil transportation markets respectively. Although each chapter is self-standing in terms of research questions, they are united by a common interest of the macroeconomic implications of the micro-level, search and matching behaviors. Specifically, chapter 2 is devoted to studying the link between the phenomenon of agricultural exodus, the time-varying patterns of labor search and matching behaviors, and their determinants in the Philippines over the past 15 years. As for chapters 3 and 4, they are dedicated to examining how the search and matching behaviors of oil tankers and exporters in the global, seaborne oil transportation industry trigger shipping market congestion and pose significant macroeconomic implications for oil trade, oil prices, and final good production. In terms of research methodology, although the chapters vary with respect to the time horizon over which they study the macroeconomic dynamics, they all explore the role of heterogeneity in shaping the aggregate outcomes, e.g., from the heterogeneity in the observed worker characteristics to that in the geographic concentration of oil trade. Given such differences and similarities of the three chapters, the overall purpose of this thesis is hence to contribute to the study of labor and oil transportation markets at a more granular level.

The process of structural transformation often goes hand-in-hand with large and persistent labor reallocations, especially from low (e.g., agriculture) to high productivity industries (e.g., service). Yet, using the Philippines Labor Force Survey (LFS) microdata, chapter 2 shows otherwise as lagging labor reallocations outside agriculture amid sustained low agricultural productivity have been a key feature in the Philippines over the past 15 years. Considering such a phenomenon, we ask what factors determine the shrinking agricultural employment in the Philippines, and in particular, which of those serve to withhold the labor market transitions from agriculture to non-agriculture.

To this end, we first show empirically that the agricultural exodus in the Philippines is a flow process from agriculture to non-agriculture, but with an intermediate step where workers are either unemployed or not-in-labor-force. Specifically, the job separation rate in the agricultural sector has been on an increasing trend over the last 15 years, whilst the agricultural job finding rate has remained relatively constant. In contrast, the job separation rate in the non-agricultural sector has remained stable, whereas the non-agricultural job finding rate has soared. This observation suggests that the shrinking agricultural employment in the Philippines can be largely attributed to increasing job separation in agriculture and greater job finding in non-agriculture.

Next, to explore the determinants underlying these labor market trends, we take a novel approach and construct a tractable equilibrium labor search and matching model featuring agriculture and non-agriculture, so as to study how changes in various economic and institutional factors interact with the worker's labor market decisions, including job separation and creation. We show analytically that at the partial equilibrium of the agricultural labor market, the worker's behaviors are captured by the interactions between a positively-sloped job separation curve and a negatively-sloped job creation curve. We further illustrate through graphical representation that several 'push' and 'pull' factors – those reflecting the incentives for workers to leave or remain in agriculture – result in shifts of

the two curves, hence leading to variations in the job separation and finding rates in the agricultural industry.

To complement the theoretical predictions and study in detail the labor market transitions in the Philippines, we propose and analyze whether some specific 'push' and 'pull' factors are behind the observed evolution of the job separation and finding rates. Through a regression analysis, we find that the widening of wage differentials with non-agricultural sectors, improvements in labor market efficiency, and better transport infrastructure are largely associated with growing outflows of labor from agriculture, whilst the lack of post-primary education and the presence of agricultural clusters hinder such outflows. Furthermore, in contrast to the traditional view that agricultural employment outflows are largely driven by productivity differences (as reflected by wage differentials), our results emphasize the roles of education as well as transport infrastructure in facilitating labor reallocations from agriculture to non-agriculture.

Overall, chapter 2 directly addresses the theme of this thesis. It considers how changes in the micro-level, labor market behaviors of the Filipino workers result in the macroeconomic phenomenon of agricultural exodus in the Philippines. Furthermore, the work assesses the quantitative importance of the observed worker heterogeneity for the dynamics in their labor market transitions; the heterogeneity in the worker's education status and ability to take advantage of better transport infrastructure is particularly emphasized such that the targeted economic policies, such as promoting wider and better education and upscaling infrastructure investment, could be formed to assist employment outflows from agriculture to non-agriculture in the Philippines.

In chapters 3 and 4, we shift our focus away from the agricultural labor market in the Philippines to the often-overlooked, global, seaborne transportation market. Indeed, despite the vital importance of seaborne transportation in championing world trade and macroeconomy, little is known about what it is comprised of, how it operates, and more importantly, why its functioning following trade shocks may lead to intended or unintended macroeconomic consequences. In particular, among the variety of commodities traded worldwide, the trade of oil – the lifeblood

of industrialized nations – is still largely dependent on seaborne transportation.[1] From the 1973 oil crisis to the 2020 Russia-Saudi oil price war, we have witnessed how tremendous and wide-reaching the impacts of such oil shocks could be to the world economy, therefore it remains critical to understand the role that the oil transportation industry plays in the shocks propagation: is it helping or aggravating the situation?

To answer this question, we first need a shipping dataset which is longitudinal, high frequency, and tracks each oil tanker's shipping status. Chapter 3 is hence developed for this very purpose. In this work, we apply the Automatic Identification System (AIS) data of the entire fleet of product tankers as they provide unprecedented details of the real-time movements of these tankers across the globe. Leveraging the information in the AIS regarding each vessel's speed, position, and draught, we first recover the loaded trips taken by the oil tankers during 2017-2020.[2] Next, to track retrospectively the searching behaviors of tankers within and across oil markets, we propose a unique procedure to recover the type of oil product the ship is carrying in each of its loaded trips, i.e., crude or product oil; the first step involves matching the exiting fixtures with known cargo types to those corresponding loaded trips identified in the AIS, whilst the second step entails an application of the Random Forest (RF) prediction model (Breiman, 2001) to predict the cargo type for those remaining, unclassified loaded trips. With this procedure, we develop a unique dataset, from which we calculate various key statistics pertaining to the shipping markets (e.g., contracting probabilities for ships, tanker utilization, etc.).

Chapter 4 applies such unique, granular shipping data to study the dynamic interactions between oil tankers and exporters following oil trade shocks that determine congestion in shipping markets. As we demonstrate, congestion is critical

[1] Huge quantities of crude oil and refined petroleum products are transported by sea between production sites, refineries, and points of consumption due to the limited pipeline infrastructure between trading partners.

[2] The draught measures the vertical distance from the bottom of a vessel's keel to the water's surface. It tells how deep the ship is sitting in the water, and reflects how much cargo the ship is carrying. A high and/or increasing draught indicates that a large and/or increasing proportion of the tanker's hull is becoming submerged, which implies that it is either loaded or in the process of loading oil.

to the macro aggregates as it entails a trade-off between the volatility of the oil trade and final good production, and that of oil prices.

We start by establishing three novel facts regarding the observed behaviors of oil tankers in the global oil transportation industry, which themselves provide the basis for the theoretical model that follows. The first is related to the presence of shipping market congestion, which we show by estimating an aggregate matching function and finding that, for a one percent increase in the number of searching ships, the average contracting probability for ships decreases significantly by 0.6 percent, holding all else constant. The second finding is associated with the prevalence of market switching behavior by tankers. By calculating their market switching probabilities, we find that, on a weekly average basis, more than one third (one fifth) of unloaded ships in the crude (product) oil market that find oil cargoes to deliver are hired by product (crude) oil exporters. The third fact is linked to the propagation of oil trade shocks in the global economy through the transportation network. We show this process by linking the observed variations in tanker utilization in each oil shipping market (i.e., the proportion of tankers that are loaded with oil cargoes in each market over the entire fleet of product tankers) to several major oil trade disruptions from 2017-2020, and find that they accord well with each other. These three facts are important, as they reflect that the shipping markets are congested, connected, and responding to oil trade shocks in a network structure.

Subsequently, we develop a Diamond-Mortensen-Pissarides (DMP) random search model that accounts for these three novel facts and incorporates the interactions between oil exporters, tankers, and a final good producer. Specifically, the oil transportation industry comprises two markets which trade crude and product oil respectively. In each shipping market, unmatched exporters have oil cargoes to deliver (which are subject to oil trade shocks) and hence participate in searching for available ships. Unloaded ships, on the other hand, constantly search for oil cargoes to deliver. The two parties engage in a random matching process with each other. If matched, unloaded ships become loaded with oil cargoes and start sailing, whilst the unmatched exporter pool is replenished with

entries of potential exporters. If unmatched, unloaded ships decide whether to continue to search in its current market or to direct their search effort to the other market, whilst unmatched exporters remain unmatched and continue searching for available ships next period. Importantly, if a ship chooses to switch markets, it must spend extra time in remaining unloaded before it becomes available in the other shipping market. This state, which we refer to as ballasting, embodies the time cost that a ship incurs when making a switching decision, since such a decision often involves the process of spatial reallocation and tank cleaning, both of which may take substantial time in practice.[3,4] Finally, after loaded ships arrive at their destinations, the transported oil cargoes are aggregated through a CES aggregator to produce a final consumption good.

With the model, we derive the equilibrium oil trade, oil prices, as well as the resulting final good production. In particular, we show that oil trade depends on the oil cargo delivered by each ship, together with the number of ships transporting oil cargoes; the former is directly determined by exogenous oil trade shocks, whilst the latter marks the corresponding endogenous adjustment of the oil transportation network to exogenous oil trade shocks. For oil prices, they are driven by the behaviors of ships across shipping markets, whereas the behaviors themselves are influenced by the imbalance between the demand for, and the supply of, oil shipping services in each market as a result of the dispersion of realized trade shocks and

[3]Using the shipping data, we find that there are significant variations in the geographic concentration of the oil trade. On the one hand, most crude oil trips originate from either the Persian Gulf or the Caribbean, and conclude in one of the major oil-importing countries in Europe (such as the U.K.) or Southeast Asia (Singapore, for example), because these oil-importing countries possess adequate oil refinery capacity. On the other hand, product oil trips (in particular, clean product oil trips) often originate from the Persian Gulf, Europe, or East Asia, but can end at almost any coastal, oil-importing country (except for those in South America and Africa). In addition, as is widely accepted in the maritime industry, whether a trip originates from east or west of the Suez Canal has significant predictive power in determining a vessel's cargo type: If the trip originates from west of the Suez Canal, there is a high probability that the ship is carrying crude oil, and vice versa. Such variations in the geographic concentration of the oil trade define the spatial nature of this industry.

[4]By industrial practice, the ship's cargo tank needs to be cleaned before it loads another type of oil product in order to avoid the risk of cargo admixture contamination. The time it takes to clean the cargo tank may vary, depending on the specific operations needed.

variations in tastes, among other primitives. Lastly, final good production combines the oil inputs and thus inherits the variations therein.

After tailoring the empirical implementation of the model to match the key statistics generated from the shipping data, we use the empirically congruent model to investigate three important and related questions: (1) How does shipping market congestion within the oil transportation network adjust to exogenous oil trade shocks? (2) How does congestion affect the searching behaviors of ships within and between markets? (3) What are the macroeconomic implications of congestion for the oil trade, oil prices, and final good production? We study the first two questions by simulating the model and show that a positive oil trade shock favors a ship's search for a contract within the affected market, hence leading to a rise in tanker utilization. As more ships switch to the impacted market, however, the market becomes less tight until the supply of shipping services exceeds demand. As a result, the benefits of the shock (greater contracting probability and commissions) start to be depleted and shipping market congestion intensifies. The congestion then drives down the contracting probability for ships and paid commissions, leading to reductions in tanker utilization. In this way, excessive flows of ships into a market that receives a favorable trade shock are prevented, such that the fluctuations in tanker utilization are attenuated.

Subsequently, we illustrate the significant implications of shipping market congestion for the trade and price of oil and for the production of a final good through counterfactual exercises. First, by comparing our baseline setup to one in which all the tankers are utilized to carry oil cargoes (hence congestion no longer manifests), we show that allowing for congestion is crucial, as it stabilizes more than one third of the volatility in the oil trade and in the final good production that would have occurred otherwise. The stabilization also comes with a more-than-tenfold increase in the volatility of oil prices, as they are directly determined by the dynamics of the shipping markets. Second, we consider another counterfactual exercise in which ships are not allowed to switch between markets, and show that the market switching behavior of ships enables asymmetric changes in congestion

between shipping markets, with distinct implications for oil trade stabilization in each market. Lastly, by contrasting our baseline setup with one in which oil trade shocks no longer occur in shipping markets, we illustrate that it is the exogenous oil trade shocks that create most of the observed aggregate fluctuations, while the corresponding endogenous adjustment of the oil transportation network and the resulting congestion effect only play a minimal role in the process.

To the best of our knowledge, chapters 3 and 4 are the first in literature to unveil the irreplaceable role played by the oil transportation industry in the propagation of oil trade shocks to the macroeconomy. Consistent with the theme of this thesis, the study explores the role of a wide range of heterogeneity (e.g., geographic dispersion of oil trade, variation in trade shocks, and idiosyncratic tastes, etc.) in contributing to the macroeconomic implications of shipping market congestion.

In what follows, the three chapters are presented consecutively as self-contained, journal-style articles, each of which contains a literature review on the relevant subject. Chapter 2 is titled "The Agricultural Exodus in the Philippines: Are Wage Differentials Driving the Process?", chapter 3 is titled "A Big Data Approach to Cargo Type Prediction and Its Implications for Oil Trade Estimation", and chapter 4 is titled "The Congestion Effect of Oil Transportation and Its Trade Implications". Chapter 5 concludes.

2

The Agricultural Exodus in the Philippines: Are Wage Differentials Driving the Process?

2.1 Introduction

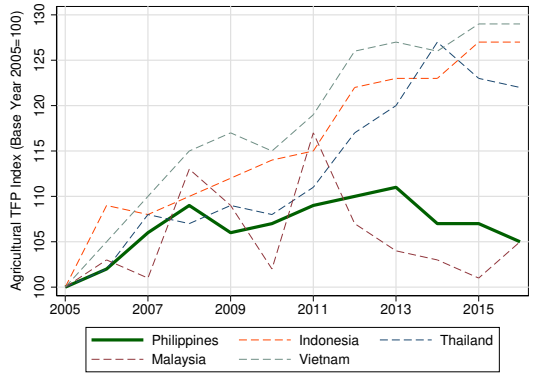
Sustained low productivity in the agricultural sector as well as lagging labor reallocations have been at the center of discussion for decades in the Philippines (e.g., Qian et al. (2018), Briones (2017b), and Teruel and Kuroda (2005)). As illustrated in Figures 2.1a and 2.1b, the Philippines' agricultural sector has underperformed its regional peers in terms of productivity growth, even though its agricultural employment has been shrinking at a similar pace to the other countries over the past 15 years. Conventional wisdom (e.g., Felipe et al. (2014) for evidence in China, McCaig and Pavcnik (2013) for evidence in Vietnam, and McMillan and Rodrik (2011) for evidence in Asia, Latin America, and Africa) suggests that low agricultural productivity translates to large wage differentials with other sectors, and drives the employment outflows from agriculture to non-agriculture.[1] Yet, studies have also shown that other social and economic factors (e.g., infrastructure as in Llanto (2012), land reform as in Adamopoulos and Restuccia (2020) and

[1]This proposition can also be found in the structural transformation literature, for instance, Herrendorf et al. (2014).

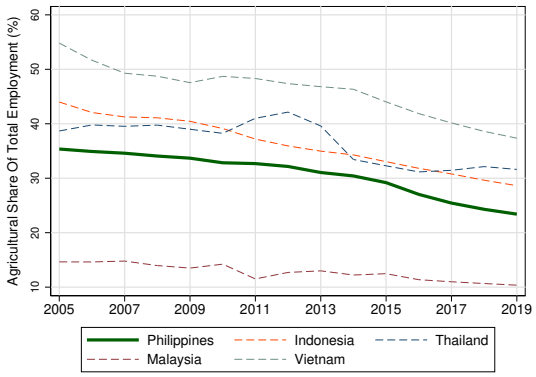
Koirala et al. (2016)) influence the agricultural worker's labor market behaviors, and therefore determine the pace of structural transformation. In light of these arguments, we aim to answer the following question: *What is driving the shrinkage of agricultural employment in the Philippines? Wage differentials or other factors?*

Figure 2.1: Agricultural performance in selected ASEAN countries

(a) Agricultural TFP



(b) Agricultural employment



Notes: (1) for agricultural TFP, the data source is 2005-2016 Agricultural Total Factor Productivity Growth Indices for Individual Countries, US Department of Agriculture. For description of methodology, see Fuglie (2015) and Fuglie (2012); (2) for agricultural employment shares, the data source is the ILOSTAT database.

To understand the evolution that takes place in the Philippines' labor market, we take advantage of the Philippines Labor Force Survey (LFS) and show empirically that the agricultural exodus in the Philippines is a flow process from agriculture to non-agriculture, but with an intermediate step where workers are either unemployed or non-participating. In this context, we find that the job separation rate in the agricultural sector has been on an increasing trend over the last 15 years, whilst the agricultural job finding rate has remained relatively constant. In contrast, the job separation rate in the non-agricultural sector has remained stable, whereas the non-agricultural job finding rate has soared. These dynamics suggest that the shrinking agricultural employment in the Philippines can be largely attributed to increasing job separation in agriculture and greater job finding in non-agriculture. Following this process across 17 administrative regions in the Philippines during 2005 to 2019, we take a novel two-step approach. Firstly, we construct a tractable equilibrium

labor search and matching model featuring two types of industries, agriculture and non-agriculture, in order to study how changes in economic and institutional factors interact with various labor market decisions, including job separation and creation. We show analytically that, at the partial equilibrium of the agricultural labor market, the worker's behaviors are captured by the interactions between a positively-sloped job separation curve and a negatively-sloped job creation curve. We further illustrate that 'push' and 'pull' factors – those reflecting the incentives for workers to leave or remain in agriculture respectively – result in shifts of the job separation and creation curves, and thus lead to variations in the job separation and finding rates in the agricultural industry. In the next step, following the model predictions which serve to discipline our empirical analysis, we propose and analyze whether some key available 'push' and 'pull' factors are behind the observed evolution of the job separation and finding rates in the Philippines. In addition to real wage differentials across sectors, we gather information on labor market efficiency, both national and regional road density as a proxy for the evolution of infrastructure across time and regions, individual's high school and/or college education experience as a proxy for human capital development, and each administrative region's share in total palay and corn production as a proxy for agricultural clusters.

By regressing each of the labor market transition rates on the proposed factors, we find that they statistically help to explain the observed variations in the job separation and finding rates for agricultural and non-agricultural workers. We then show that sustained low productivity in agriculture translates to widening wage differentials, and thus seems to facilitate agricultural workers to separate more from their current jobs and direct their search effort elsewhere; such motives become especially prominent when labor market efficiency is improving in the non-agricultural sectors and transport infrastructure is upgraded, as agricultural workers can switch industries more freely. On the other hand, in terms of the effects from the 'pull' factors, the lack of sufficient education limits the ability of agricultural workers to transition into non-agricultural jobs, and essentially drags them back to agricultural employment due to fewer outside options; at

the same time, the difficulty of switching out of the agricultural sector is also heightened when the region's economy is more dominated by agriculture. These results are statistically robust to considering cross-region real wage differentials, job-to-job transitions, as well as job separations and findings in the low-skilled non-agricultural industries, among others.

The results support that not only do increasing wage differentials affect the employment outflows from agriculture to non-agriculture, but other economic and institutional factors also play an important role in such transitions. Moreover, we find that contrary to the traditional view, in spite of the statistical significance, increasing real wage differentials do not contribute much to the agricultural exodus in the Philippines. Instead, measures related to enhancing the worker's ability to freely switch industries are both statistically and economically significant in explaining the observed developments in labor market transition rates. Specifically, high school education can add significantly more to the exodus through raising agricultural worker's job separation margin substantially.[2] Building on that, when college education is undertaken, employment outflows from agriculture to non-agriculture could be further amplified because workers can now find and hold on to non-agricultural jobs more easily. Improving transport also plays a more economically significant role in driving the agricultural exodus, as not only does higher road density reduce job finding in agriculture by a wide margin, but also it boosts job finding in non-agriculture significantly.

This paper makes two contributions to the literature. To the best of our knowledge, this paper is the first to estimate time series trends of the labor market transition rates in the Philippines and simultaneously examine the determinants through which changes in the transition rates give rise to shrinking employment in agriculture. Previous research largely focuses on whether a certain economic or demographic factor affects agricultural employment directly. We argue that this approach is too holistic as it ignores the important labor market transitions between the pools of employment and non-employment which underlie the employment

[2]This is the case even when we compare it to the effect on job separation margin in low-skilled non-agricultural industries. See the third robustness check in Section 2.3.3 for details.

outflows from agriculture to non-agriculture. In addition, this paper offers an assessment on the importance of each identified 'push' or 'pull' factor as a potential driving force of the agricultural exodus. We apply both the empirical assessment of economic significance and structural approach (embedded in the theoretical model) to illustrate the vital importance of providing channels/means to agricultural workers to facilitate their inter-sectoral reallocations.

These considerations are important not only for improving the performance of the agricultural sector in the Philippines, but also for lowering the scarring effects of the COVID-19 shock through more efficient inter-sectoral labor reallocations. Our results highlight the need to maintain efficient labor markets and increase investment in education and transport infrastructure, as a complement to finishing land reform measures that cannot be captured directly in our empirical study. As agriculture weakens and economies develop, a well-functioning labor market is crucial in reducing rural poverty and rural-urban income disparities, and policies can be developed in this regard to ensure effective preparedness of those who choose to leave agriculture. In addition, since a wider coverage of high school education proves sufficient to speed up labor reallocations from agriculture to non-agriculture, policymakers should pay more attention to promoting education and upgrading the skills of agricultural workers, especially those residing in regions with low education attainment and that could potentially benefit the most from education (e.g., the young and female). In addition, infrastructure investment could be further expanded since it also assists employment outflows from agriculture to non-agriculture.

The rest of this paper is organized as follows. Section 2.2 is devoted to establishing the facts of the labor market in the Philippines, including time series trends of the job separation and finding rates. Section 2.3 presents the baseline empirical results, identifies the relevant 'push' and 'pull' factors in explaining the observed developments in labor market transition rates, and performs robustness checks to support our main empirical findings. Section 2.4 studies the relative importance of each identified factor and details our policy suggestions. Section

2.5 concludes. The Appendices contain additional model details, analytical results, robustness checks, and additional figures and tables.

2.2 Labor Market in the Philippines

In this section, we document three aspects of the Philippines' labor market: firstly, we present the Labor Force Survey data; secondly, taking a stock-flow approach, we estimate time series trends of the job separation rate (or Employment to Non-employment rate, EN) and job finding rate (or Non-employment to Employment rate, NE) in both agricultural and non-agricultural sectors; and finally, we propose a theoretical framework and identify the evolution of some empirically available 'push' and 'pull' factors in affecting the agricultural worker's labor market behaviors.

2.2.1 Data

One of the main objectives of this paper is to examine the underlying reasons for the shift in employment from the agricultural to non-agricultural sectors. This is possible by using panel data that track each worker's labor market status. Therefore, we apply the Philippines Labor Force Survey (LFS) Microdata Public Use Files. The LFS tracks a nationally-representative sample of the Philippines' inhabitants, and the cleaned dataset contains in total nearly 11.4 million quarter-individual observations, spanning from 2005Q2 to 2019Q4 and covering 17 administrative regions. Since its initiation in the 1980s, the LFS has undergone several structural changes and the definition of employment/unemployment/non-participation varied significantly across different years. Considering this, we decide to restrict the sample from 2005Q2 to 2019Q4 because (1) the identification of discouraged workers is only available from 2005Q2 onward when the inquiries that were deemed necessary to adequately capture the availability criterion and to reflect the reference period started to be incorporated; and (2) the questions related to worker's employment in the past quarter are only available from 2005Q2 onward; these questions are critical in the sense that they allow us to capture changes in employment status and industry in which the worker is employed. The dataset contains in total 21

variables, ranging from the worker's basic demographic characteristics to his/her employment status and daily pay.

We use the LFS because of its three unique features: it is (1) a longitudinal dataset with (2) quarterly data that (3) records each individual's labor market behaviors. These three features are necessary to identify what factors are driving the employment outflows from agriculture to non-agriculture. We need labor market panel data in order to observe the outcomes of the worker's job search, and the data also need to be of sufficiently high frequency in order for us to construct statistically meaningful job separation and finding rates and to avoid the time aggregation bias (e.g., Shimer (2012), Elsby et al. (2009), and references therein). However, despite the richness and usefulness of information in the LFS, it is a micro panel with large N but small T: even though we observe the changes in labor market behaviors by individuals between two adjacent quarters, we are unable to uniquely identify each individual across quarters due to the lack of individual and geographic ID variables. In addition, the survey question in the LFS asking each worker's employment status in the past quarter does not distinguish unemployment from non-participation. This leads us to group the unemployed with non-participants and study the labor market behaviors by the non-employed in general. This limitation as unemployment and non-participation are distinct labor market statuses over the business cycles is not critical in emerging economies. Specifically, Elsby and Shapiro (2012) and Juhn et al. (2002), Juhn et al. (1991) argue that the boundary is blurred over the long-run. At low frequencies, the unemployed resemble non-participants because they have relatively long spells of joblessness and minimal employment opportunities. Moreover, Bosch (2016) illustrates that the lack of unemployment insurance in emerging economies gives rise to workers who experience long spells of unemployment losing momentum to continue to search, essentially behaving as non-participants in the long-run.

We restrict our sample to individuals with age of 15 to 64, and employed workers excluding public administration, defense, compulsory social security, and extra-

territorial organizations and bodies.[3] We exclude individuals below the age of 15 because according to the Labor Code of the Philippines Implementing Rules and Regulations, even though the minimum age for employment is 18 years old, persons of age 15 to 18 can be employed given that they work in non-hazardous environments. We also keep individuals above the age of 60 but below 65 even though the official retirement age in the Philippines is set at 60. This is because the retirement age for an employee depends specifically on the employment contract. Only if there were no existing retirement plan or agreement for the employee might he/she retire at the age of 60, given that he/she has served the employer for 5 years, and shall be given a retirement pay of at least half a month's salary for every year of service (6 months of work given is considered as 1 whole year for the retirement pay). In addition, by examining the LFS data, we find that there are on average 3.8% of employed workers each quarter that have an age between 60 and 64, therefore representing a non-negligible share of employment in the Philippines. Lastly, we exclude individuals working in public administration, defense, compulsory social security, and extra-territorial organizations and bodies so as to restrict our attention to the employment/non-employment dynamics in the private sector labor market.

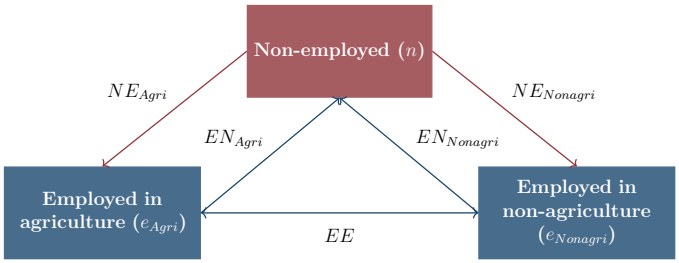
2.2.2 Time Series Trends of the EN and NE Rates

Changes in employment are often driven by variations in the labor market transition rates. Therefore, in order to understand why the agricultural employment share has been declining in the Philippines during the past 15 years, we dissect the labor market and take a stock-flow approach. As illustrated in Figure 2.2, we see that while the stock of employment in agriculture e_{Agri} is directly affected by its outflow and inflow rates, EN_{Agri} and NE_{Agri}, the transition rates in non-agriculture, $EN_{Nonagri}$ and $NE_{Nonagri}$, also have an indirect impact through altering the composition of the non-employment pool. Note that in practice there are job-to-job (EE) transitions, through which workers switch industries without experiencing a non-employment spell in between. However, we do not model or include in the empirical baseline

[3]See Appendix A.1 for the detailed industry classification.

such transitions because agricultural (non-agricultural) workers only switch to non-agricultural (agricultural) employment through job-to-job transitions with an average probability of 2.4% (0.9%) during the sample period.

Figure 2.2: Labor market composition



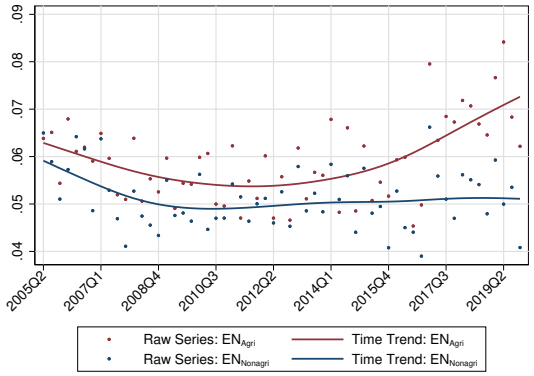
Notes: the dynamics in the Philippines' labor market can be illustrated by the flows between the stocks of non-employment (n), employment in agriculture (e_{Agri}), and employment in non-agriculture ($e_{Nonagri}$). Specifically, EN_{Agri} and $EN_{Nonagri}$ represent the outflow rates from agricultural and non-agricultural employment respectively, NE_{Agri} and $NE_{Nonagri}$ represent the inflow rates into agricultural and non-agricultural employment respectively, whilst EE represents the job-to-job transition rate between agricultural and non-agricultural employment.

After establishing the links between agricultural employment and the corresponding labor market transition rates, we define the samples used to calculate the EN and NE rates. For the EN rates, the underlying sample contains matched records that start with employed in the past quarter; the dependent variable takes 1 when an EN transition occurs and 0 when the worker remains employed. Likewise, the sample for NE rates consists of those who are non-employed in the past quarter; the variable takes 1 if the worker is employed in the survey quarter and 0 otherwise. To calculate the EN rates (NE rates) for agricultural and non-agricultural workers respectively, we further classify each sample based on whether the worker is employed in the agricultural sector before non-employment (after non-employment).

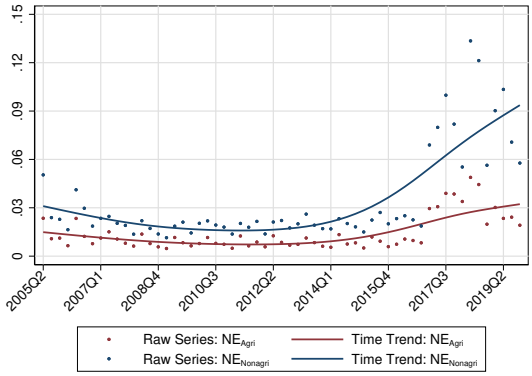
Lastly, to visualize the long-run trend of each labor market transition rate, we firstly calculate the weighted average EN and NE rates based on the final person weights provided for each survey quarter, and then apply the Hodrick-Prescott Filter to extract the trend component of each series; the smoothing parameter is set to 1600.

Figure 2.3: Time series trends of the labor market transition rates

(a) EN rates



(b) NE rates



Source: Labor Force Survey, Philippine Statistics Authority. The trends are estimated using the Hodrick-Prescott Filter with smoothing parameter of 1600.

2.2.2.1 EN Rates

Figure 2.3a plots time trends and unadjusted raw series of the EN rates for agricultural and non-agricultural workers respectively. The unadjusted raw series are average quarterly means, weighted by the LFS final person weights. By comparing the trends, we can clearly see that in levels, the EN rate is on average higher for agricultural workers, and the difference is widening in more recent years. In the early sample, both the EN rates for agricultural and non-agricultural workers have been declining, while in the latter half, the EN rate for agricultural workers rises sharply and overtakes its 2005 level, reaching 7.3% at the end of the sample; during the same time, the EN rate for non-agricultural workers remains relatively stable.

2.2.2.2 NE Rates

Figure 2.3b plots time trends and unadjusted raw series of the NE rates in agriculture and non-agriculture respectively. The NE rate is on average lower in agriculture, and the difference is also increasing more recently. Both NE rates remain low before 2016 when they start climbing; the rise of NE rate in the non-agricultural sector is significantly larger than the one in agriculture.

In summary, by examining time series trends of the EN and NE rates, we conclude that the agricultural exodus in the Philippines is a flow process from agriculture to non-agriculture, but with an intermediate step where workers are either unemployed or non-participating; it is mostly driven by the increasing outflows from agricultural employment to non-employment as well as larger inflows from non-employment to non-agricultural employment.

2.2.3 'Push' and 'Pull' Drivers of Intersectoral Labor Reallocations

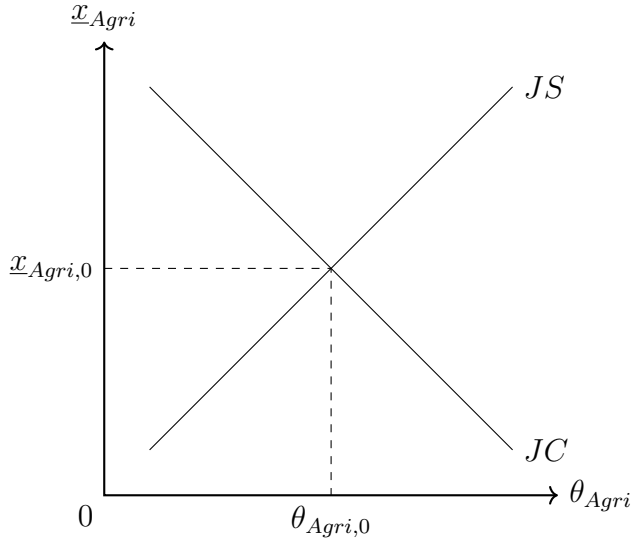
In the previous subsection we show that there exist drastic differences in the labor market transition rates for agricultural and non-agricultural workers, and the disparity has widened in recent years. In order to understand how the EN and NE rates are shaped, and explore the extent to which the social and economic reforms in the Philippines affect an individual's job separation and finding probabilities, we follow the theoretical foundation and propose within its discipline two specific sets of factors, with 'push' factors capturing incentives for workers to leave agriculture and 'pull' factors capturing incentives for workers to remain in agriculture. Subsequently, we assess the importance of each proposed factor in the context of the Philippines.

2.2.3.1 Theoretical Foundation

We start with a theoretical approach by constructing a model of the labor market. Then, we use the model to study the influence of 'push' and 'pull' factors in general on job separation and finding in the agricultural sector in a systematic manner. Specifically, to capture the empirical observation that job openings and job seekers simultaneously exist, we build a two-market random search model in the spirit of the seminal work by Mortensen and Pissarides (1994), where labor market frictions prevent job openings and job seekers from perfectly matching up. The model accounts for the job finding, separation, and industry switching behaviors by workers in both agriculture and non-agriculture while remaining analytically tractable.[4]

[4]Details of the model can be accessed in Appendix A.2.

Figure 2.4: Job separation and creation in the agricultural labor market



Notes: (1) $\underline{x}_{Agri}, \theta_{Agri}$ stand for reservation productivity and labor market tightness in the agricultural labor market respectively; (2) $(\underline{x}_{Agri,0}, \theta_{Agri,0})$ depicts the partial equilibrium by the intersection of the JS and JC curves; and (3) for graphical representation, we only plot the JS and JC curves as straight lines and neglect the changes in slopes at different combinations of $(\underline{x}_{Agri}, \theta_{Agri})$.

Through the lens of the model, as in Figure 2.4, we show that there exist a positively-sloped job separation (JS) curve and a negatively-sloped job creation (JC) curve in the plane containing reservation productivity and labor market tightness in agriculture. Intuitively, reservation productivity in the labor market refers to the cutoff productivity below (above) which workers and firms choose to sever (continue) their employment relationship, while labor market tightness measures the balance between the demand for, and the supply of, labor – the tighter the labor market is, the higher the demand for labor will be. The positive slope of the JS curve captures a notion that higher demand for labor (as measured by larger θ) implies greater outside option value for workers, hence leading to more job offers which do not promise sufficiently high returns being declined.[5] In terms of the negative slope of the JC curve, it captures exactly the opposite: higher demand for labor implies smaller outside option value for firms as it becomes more

[5]In the model, when bargaining over an employment relationship, the outside option of the worker is to either remain unmatched in the current industry or to switch to the other industry, whilst the outside option of the firm is to remain unmatched and wait for another worker.

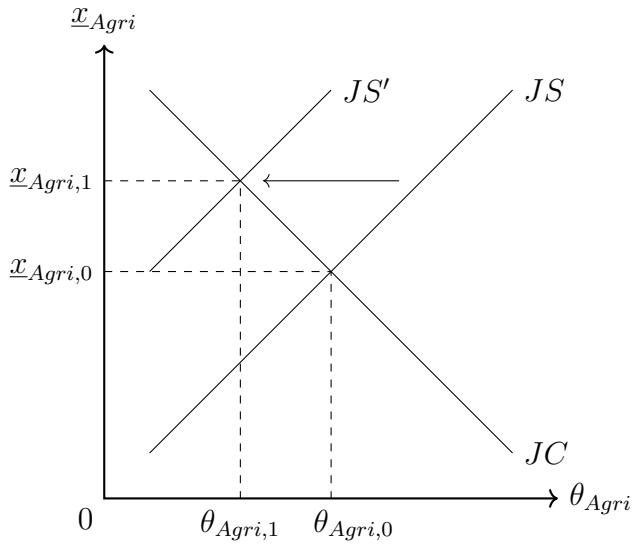
difficult to fill a vacancy. In consequence, firms lower their reservation productivity to attract more labor. Lastly, the intersection of the two curves marks the partial equilibrium in the agricultural labor market, and any shift in the JS or JC curve due to changes in labor market conditions will result in corresponding changes in the underlying endogenous variables, and thus affect the equilibrium job separation and finding rates in the agricultural sector.

In terms of the 'push' factors, let's first consider an exogenous increase in the non-agricultural productivity such that the outside option for workers switching from agriculture to non-agriculture becomes more favorable. This is captured in the model with a leftward shift of the JS curve to JS' as depicted in Figure 2.5. Assuming all else remains constant, at the new equilibrium where the JS' and JC curves intersect, the reservation productivity increases whilst the agricultural labor market tightness drops. Since the separation rate in agriculture can be shown to monotonically increase in the reservation productivity, such changes imply greater job separation in agriculture in equilibrium; since the outside option value for agricultural workers increases as a result of the higher non-agricultural productivity, they become more 'picky' about job offers and decline more often. In terms of the job finding rate, it is monotonically increasing in agricultural labor market tightness and decreasing in reservation productivity. Therefore, the confluence of higher reservation productivity and lower tightness in equilibrium implies a lower job finding rate in agriculture; not only do agricultural workers decline offers more often, but also they face more difficulties in meeting a potential employer. In reality, such a productivity difference often manifests itself through widening wage differentials, therefore incentivizing agricultural workers to move away from their land and search for non-agricultural jobs instead.

Similar dynamics can arise from an exogenous increase in the flow outside value for non-agricultural workers or an exogenous increase in the matching efficiency in the non-agricultural industry, as they both behave to raise the outside option value faced by agricultural workers.[6] It is also noted that the increasing productivity

[6]For reference, the flow outside value refers to the per-period utility that the worker gets during non-employment, e.g., unemployment benefits and leisure.

Figure 2.5: The influence of higher non-agricultural productivity on job separation and finding



Notes: (1) JS' represents the shifted job separation curve following an exogenous increase in the non-agricultural productivity; (2) the intersection between the JS' and JC curves represents the new (partial) equilibrium in the agricultural labor market, featuring $(\underline{x}_{Agri,1}, \theta_{Agri,1})$.

difference between agriculture and non-agriculture may result from lower agricultural productivity instead (recall Figure 2.1a), which in certain scenarios (see Figure A.1 for an illustration) will lead to both the JS and JC curves shifting to the left, and yield similar outcomes to those with only the JS curve shifting to the left.

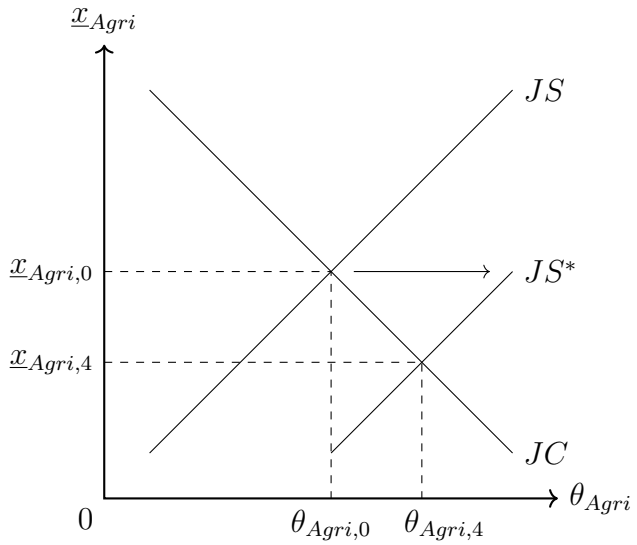
In addition to the aforementioned 'push' factors, a boost to matching efficiency in the agricultural industry may also lead to significant employment outflows from agriculture, even though the exact mechanism through which the outflows take place varies from the previous ones. Specifically, it triggers not only a leftward shift of the JS curve, but also a rightward shift of the JC curve. The intuition is that the boost to matching efficiency in the agricultural industry not only enhances the outside option for its workers (as they can find jobs more easily), but also improves the ability of firms in filling their vacancies and hence encourages them to create more agricultural jobs. Graphically, in Figure A.2, assuming that the shift in the JC curve is relatively small, at the new equilibrium where JS''' and JC''' curves intersect, the reservation productivity increases whilst the agricultural labor market tightness falls. This leads to similar outcomes as in the other cases,

namely, greater job separation but lower job finding probabilities in the agricultural industry. Therefore, given the above model predictions, it is expected that better-functioning labor markets and increasing measures to boost matching efficiency in both agriculture and non-agriculture (e.g., improved transport infrastructure) would serve to promote agricultural employment outflows – proxies of these 'push' factors will be formally tested later in the empirical section.

Now we turn to the 'pull' factors and analyze how they influence the job separation and finding rates in agriculture. We start with an exogenous decrease in the flow outside value for agricultural workers. Holding all else constant, this implies a rightward shift of the JS curve to JS^* as depicted in Figure 2.6. As a result, at the new equilibrium where JS^* and JC curves intersect, the reservation productivity declines whilst the agricultural labor market tightness rises; compared to the original equilibrium, the job separation rate falls whilst the job finding rate increases in the agricultural industry. The intuition is simple: the decrease in the agricultural worker's outside value enlarges the surplus to be shared, which then leads to lower realizations of idiosyncratic productivities – which were previously rejected – are now accepted; in other words, workers become less 'picky' and are less likely to separate. In practice, the decrease in the flow outside value for agricultural workers is often associated with a lack of education or skill depreciation, since they cannot satisfy the skill requirements for non-agricultural jobs and thus get stuck in agriculture.

Another 'pull' factor to note is the deteriorating bargaining power of agricultural workers in their wages negotiation, which may also arise in the case of obsolete skills or agricultural clusters. Yet, the mechanism through which it hinders agricultural exodus differs from that illustrated in Figure 2.6; it instead involves rightward shifts of both the JS and JC curves. The dynamics are shown in Figure A.3, and the new equilibrium is where the JS^{**} and JC^{**} curves intersect, provided that the shift in the JC curve is relatively small. The new equilibrium features lower job separation but higher job finding probabilities in the agricultural industry because on the one hand, the lower bargaining power of workers shrinks their share of the matching surplus and discourages job separation, while on the other

Figure 2.6: The influence of lower flow outside value for agricultural workers on job separation and finding



Notes: (1) JS^* represents the shifted job separation curve following an exogenous decrease in the flow outside value faced by agricultural workers; (2) the intersection between the JS^* and JC curves represents the new (partial) equilibrium in the agricultural labor market, featuring $(\underline{x}_{Agri,4}, \theta_{Agri,4})$.

hand, the resulting greater bargaining power of firms increases their share of the matching surplus and promotes job creation. Hence, given the model predictions, it is expected that inadequate education and significant agricultural clusters would serve to dampen agricultural employment outflows.

Having illustrated how the 'push' and 'pull' factors in general determine the evolution of EN (job separation) and NE (job finding) rates, we start to introduce some specific 'push' and 'pull' factors; the selection is based on both empirical relevance and data availability.

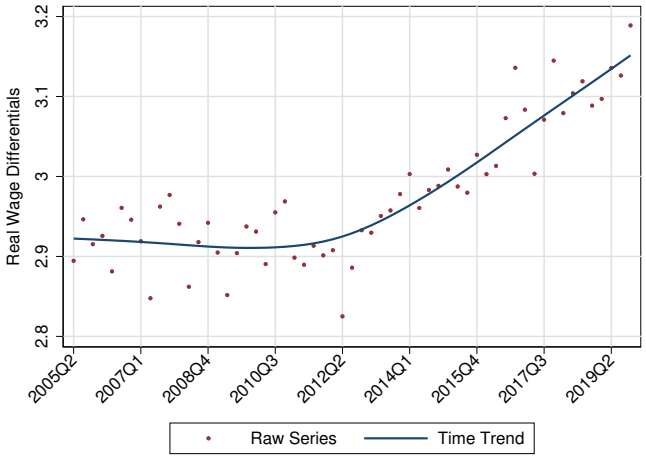
2.2.3.2 Empirically Available 'Push' Factors

Real Wage Differentials. Having sustained average annual growth of 6.4% between 2010-2019, from an average of 4.6% between 2001-2009, the Philippines' economy is considered one of the most dynamic economies in East Asia and the Pacific. However, the extent to which this growth has benefited agricultural workers, who represent a sizable share of the country's workforce (from 35% in 2005 to 24% in 2019), remains unclear. With a basic pay less than half that of a typical

Filipino worker, agricultural workers represented roughly two of every three working poor in the country in 2012 (Briones, 2017a).

More strikingly, agricultural productivity in the Philippines is already lagging behind the average. As illustrated in Figure 2.1a, the agricultural TFP growth for the Philippines continued to decline during 2001-2016, even diving into a negative trajectory from 2013 onward. Compared to other ASEAN countries, the growth momentum in the Philippines' agricultural sector remains lackluster and far below the region's average. Such sustained low productivity – resulting from historical structural inefficiencies – is translated into large wage differentials (see Figure 2.7) between the agricultural and non-agricultural sectors, and these differences (which could be more severe by province) would induce workers to switch out of agriculture.

Figure 2.7: Real wage differentials



Source: Labor Force Survey, Philippine Statistics Authority. Real wage differentials are calculated as the natural logarithm of the differences in weighted average real hourly wages between non-agricultural and agricultural workers in each region, i.e., $w_{x,Nonagri} - w_{x,Agri}$ for region x. The weight applied to each individual is his/her share in the total employment in the agricultural/non-agricultural sector during the survey quarter, which is defined as the sum of the products of each agricultural/non-agricultural worker's weekly working hours and his/her final person weight in that quarter. The time series trend is estimated using the Hodrick-Prescott Filter with smoothing parameter of 1600.

Labor Market Efficiency. In the absence of direct measures on labor market efficiency, we draw on the information from World Economic Forum Global Competitiveness Index (Schwab, 2019) in order to gauge how efficiently the Philippines' labor market has evolved. The Global Competitiveness Index builds on the Executive

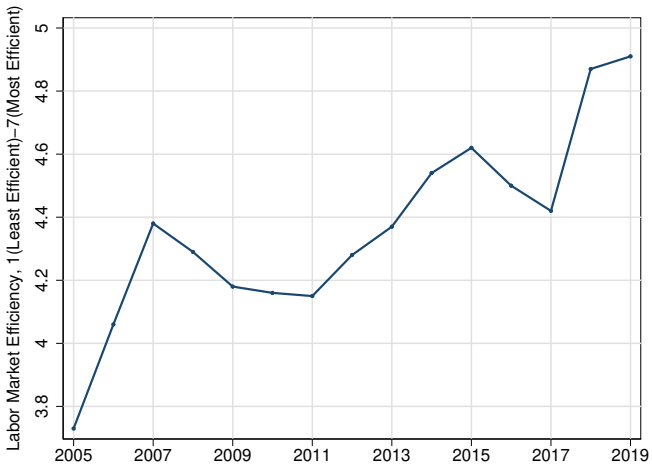
Opinion Survey. For the last 40 years, it has been a key ingredient of the index, and captured the views of surveyed business executives in member countries. Most questions in the survey ask respondents to evaluate on a scale of 1 (considered among the worst in the world) to 7 (considered among the best in the world) the performance on various topics of the country where the respondent operates. The sample frame reflects the underlying structure of the economy as it is (1) in proportion to the share of GDP by sector, (2) ensures the representation of both large- (more than 250 employees) and small-sized companies (250 employees or fewer), and (3) ensures that the chosen companies also have a sufficiently wide geographic coverage. For our purpose, we calculate the arithmetic average of the following five indices to construct the labor market efficiency index: flexibility of hiring and firing practices, cooperation in labor-employer relations, flexibility of wage determination, reliance on professional management, and the extent to which pay reflects the employee's productivity.[7] In addition to data availability, we choose these indices following the literature that flexible contractual arrangement (e.g., International Monetary Fund (1999), Organisation for Economic Co-operation and Development (1994)) and efficient use of talents (e.g., Ostoj (2015)) are critical for the efficient functioning of labor market.

As captured in Figure 2.8, despite some fluctuations, the index is on an increasing trend during the sample period. This implies improving employment absorption in the non-agricultural sector, which in theory reduces labor market frictions and strengthens the worker's ability to switch industries. As a result, the increasing labor market efficiency accompanies rising wage differentials and leads to significant employment outflows from agriculture to non-agriculture.

Road Density. Similar to the role of labor market efficiency, we propose that the enhancement in road infrastructure, as measured by increasing road density, leads to lower reallocation costs and facilitates transitions out of the agricultural sector.

[7]As the sample frame aims to replicate an economy's sectoral composition and includes companies of different sizes, the country-level score of each question is the arithmetic mean of all answers in each country. That is, for a given question, all individual answers carry the same weight.

Figure 2.8: Labor market efficiency index



Source: Global Competitiveness Index, World Economic Forum.

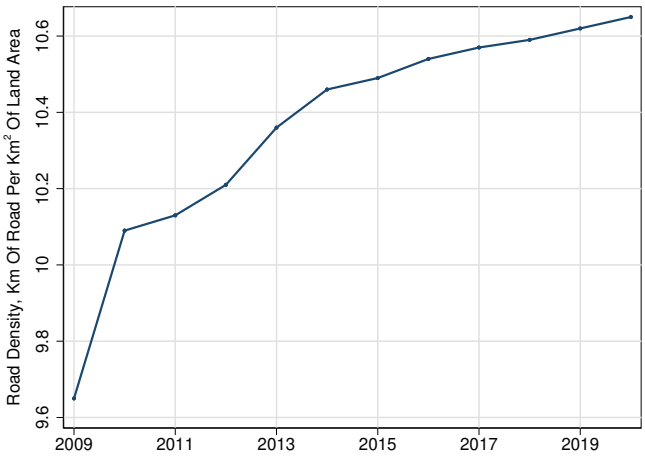
Applying the information in 2018 and 2019 ATLAS published by the Department of Public Works and Highways (DPWH), we show in Figure 2.9 that national road density continued to intensify during the last ten years, despite the rapid pace of land developments in the Philippines. This can lead to better connectivity in agriculture and rural development, as it has been suggested that agricultural workers and those living in sparsely populated areas are effectively handicapped in mobility terms to access services, carry out everyday activities or maintain social links (European Commission, 2017).

Notwithstanding the overall increase, there is still wide disparity in road density across regions. In 2019, National Capital Region (NCR) has 188.24 kilometers of roads per square kilometer of land area whereas Western Visayas, the region with the second highest road density, has only 23.91 kilometer per square kilometer. Cagayan Valley has the lowest road density among all, at 6.73 kilometers per square kilometer.

2.2.3.3 Empirically Available 'Pull' Factors

High School and College Education. The importance of education to labor reallocations is stressed in the literature that integrate migration and human capital theory, of which Sjaastad (1962) provided one of the first examples. This strand of literature (e.g., Maucorps et al. (2019), Tocco et al. (2013), and Bojnec and

Figure 2.9: National road density

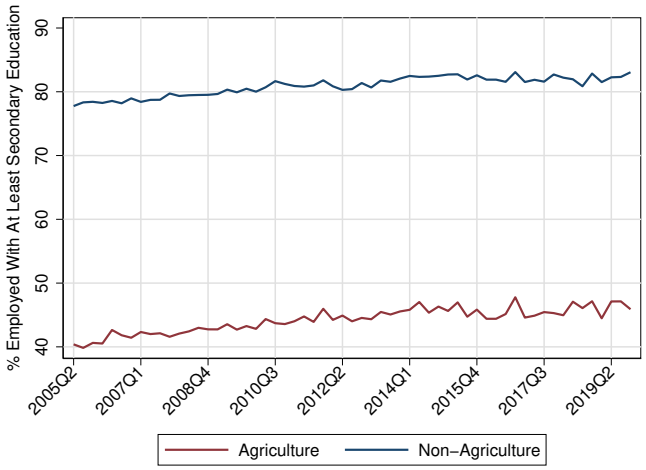


Source: 2018 & 2019 ATLAS, Department of Public Works and Highways. Road density pertains to the ratio of the length of the country's total road network to the country's land area. The road network includes all roads in the country: motorways, highways, main or national roads, secondary or regional roads, and other urban and rural roads.

Dries (2005)) predicts that better-educated individuals or those with specific social capital (e.g., networks and contacts beneficial for intersectoral reallocations) can have a strong advantage over others because better access to information and low transaction costs are critical, especially when changing jobs also requires switching sectors. Furthermore, studies have found that inadequate human capital is a critical constraint, not only for agricultural labor restructuring, but also more generally for business development and economic activities in rural areas (Rizov & Swinnen, 2004).

As such, we gather information on both high school and college education attainment in the Philippines as a proxy for human capital development. We show in Figure 2.10 that, there is a significant difference in higher education attainment between agricultural and non-agricultural workers in the Philippines, with an average of over 80% of employed workers in non-agriculture but less than 50% in agriculture reaching at least high school education. This lower level of education constrains agricultural workers from seeking employment in more skill-demanding industries, and raises the costs they will potentially incur in switching from the agricultural to non-agricultural sectors (e.g., job-specific training).

Figure 2.10: Share of employed workers with at least high school education



Source: Labor Force Survey, Philippine Statistics Authority.

Palay and Corn Production. We also examine the effects of industrial clusters on employment outflows from agriculture. Brenner and Gildner (2006) find that industrial concentration may have a negative impact on the local economic situation and resource allocation due to the lock-ins. For instance, if agriculture develops very well in a region, it might hinder other industries from also flowering there. This argument can be firstly based on the consideration that a successful industry with many employees increases prices for factors of production (e.g., labor) in the region. This makes the region less attractive for other industries as they are unable to benefit from economies of location in connection with the already existing industry.

Besides this simple cost argument, poverty (Carvalho et al., 2010), which is more prominent in rural areas that live off agriculture, creates psychological poverty traps, whereby people are too much focused on meeting immediate basic needs (feeding and sheltering) and forego investment opportunities due to impaired vision of the future and exorbitant time discount rates. Consequently, local agricultural clusters provide a disadvantageous ground for development in other industries and employment therein.

To measure the prevalence of agricultural clusters in an administrative region, ideally we would apply the agricultural/non-agricultural job vacancy data or the agricultural share of real value-added in each region. However, neither of them is

available for the entire sample period. Instead, we apply each region's share of total palay and corn production in the Philippines.[8] We hypothesize that, the higher the region's share is in the total palay and corn production, the more dominant the agricultural sector is in the regional economy, and hence the larger the lock-in effect will be in restraining employment outflows from agriculture to non-agriculture.

2.3 Empirical Analysis

Having introduced some key aspects of the labor market in the Philippines, we present in this section the baseline regression results.

2.3.1 Linear Probability Model

In the previous sections, we propose a non-exhaustive but empirically available list of 'push' and 'pull' factors to account for the observed variations in the labor market transition rates. In order to identify the roles of these factors, we build a Linear Probability Model (LPM) of job separation and finding. The LPM allows us to easily and systematically control for changes in demographic and regional compositions over the sample period. Furthermore, despite the shortcoming as it does not constrain predicted probabilities to the unit interval, the LPM provides a reasonable direct estimate of the sample-average marginal effect on the probability that $y = 1$ as x changes, as well as that it provides a good guide to which variables are statistically significant (Cameron & Trivedi, 2005).

Our LPM assumes that, for the binary outcome (Bernoulli trial), EN, and its associated vector of explanatory variables, X,

$$Pr(EN = 1|X = x) = x'\beta,$$

where X includes real wage differentials, labor market efficiency index, logged road density, high school education dummy, college education dummy, each region's share of total palay and corn production (all of which are interacted with region dummies),

[8]For reference, the regional share of total palay and corn production has a correlation of 0.46 with the agricultural share of real value-added in each region for 2009-2018, indicating that it is a reasonable proxy for agricultural clusters.

worker's demographic characteristics (gender, age, and martial status), and year and quarter dummies. We apply heteroskedasticity-robust standard errors.[9] The same specification is also applied to the NE rates.

Note that since the 2018 and 2019 DPWH ATLAS record road density data in 2009-2019 for all administrative regions except for Bangsamoro Autonomous Region in Muslim Mindanao (BARMM), in order to keep observations that would otherwise be dropped in the regressions, we apply the dummy variable adjustment as in Cohen et al. (2013) such that (1) BARMM is assigned the national average road density, and the road density for each region during 2005-2008 is assumed to equal their 2009 level; and (2) a dummy variable coded 1 if road density is imputed and 0 otherwise is included in the regressions.

2.3.2 Baseline Regression Results

Table 2.1 summarizes the average marginal effects (AMEs) of each of the aforementioned factors on the labor market transition rates. By examining the R-squared and F statistics associated with the joint significance of the identified 'push' and 'pull' factors, their interactions with region dummies, and demographic dummies, we find that even though the observed developments in the EN and NE rates are not so well replicated by the model, the proposed factors are jointly significant at the 1% level of significance, indicating their statistically significant explanatory power.[10]

The estimated AMEs of real wage differentials are significantly positive for the outcomes EN_{Agri} and $NE_{Nonagri}$, whilst significantly negative for the outcomes

[9]Recall that the variance of a Bernoulli random variable Y is given by $Var(Y) = Pr(Y = 1) \times (1 - Pr(Y = 1))$. Hence, in a LPM with general form $Y_i = \beta_0 + \beta_1 X_{1,i} + \cdots + \beta_k X_{k,i} + u_i$, we can write the conditional variance of the error term as:

$$\begin{aligned} Var(u_i|X_{1,i}, \ldots, X_{k,i}) &= Var(Y_i - (\beta_0 + \beta_1 X_{1,i} + \cdots + \beta_k X_{k,i})|X_{1,i}, \ldots, X_{k,i}) \\ &= Var(Y_i|X_{1,i}, \ldots, X_{k,i}) \\ &= Pr(Y_i = 1|X_{1,i}, \ldots, X_{k,i}) \times Pr(1 - Pr(Y_i = 1|X_{1,i}, \ldots, X_{k,i})) \\ &\neq \sigma_u^2. \end{aligned}$$

Hence, it is essential to use heteroskedasticity-robust standard errors in a LPM.

[10]For reference, the R-squared coefficients for the model in which we only regress each of the transition rates on region and time dummies are 0.0063 for EN_{Agri}, 0.0027 for $EN_{Nonagri}$, 0.0146 for NE_{Agri}, and 0.0260 for $NE_{Nonagri}$ respectively.

Table 2.1: AMEs in the baseline LPM

	EN_{Agri}	$EN_{Nonagri}$	NE_{Agri}	$NE_{Nonagri}$
'Push' Factors				
Real Wage Differentials	0.00947***	-0.00572***	-0.00683***	0.00631***
	(0.00245)	(0.00203)	(0.000991)	(0.00191)
Labor Market Efficiency	0.00675**	-0.0175***	-0.00174	0.0496***
	(0.00279)	(0.00289)	(0.00167)	(0.00191)
Road Density	0.0487**	-0.0170	-0.0768***	0.0552***
	(0.0213)	(0.0192)	(0.0115)	(0.0180)
'Pull' Factors				
High School Education	0.0204***	0.00276***	-0.0190***	-0.00237***
	(0.000473)	(0.000469)	(0.000342)	(0.000475)
College Education	0.0319***	-0.0118***	-0.0267***	0.00173***
	(0.000935)	(0.000482)	(0.000333)	(0.000510)
Palay & Corn Production	-0.112***	-0.0208	0.00976	-0.0340**
	(0.0164)	(0.0170)	(0.00807)	(0.0154)
Demographic Characteristics				
Female	0.0925***	0.0115***	-0.0311***	-0.0424***
	(0.000627)	(0.000343)	(0.000276)	(0.000445)
Age 25-34	-0.0541***	-0.0388***	0.00568***	0.0300***
	(0.000838)	(0.000616)	(0.000297)	(0.000586)
Age 35-44	-0.0628***	-0.0529***	0.00889***	0.0327***
	(0.000864)	(0.000643)	(0.000369)	(0.000697)
Age 45-54	-0.0627***	-0.0533***	0.00866***	0.0232***
	(0.000890)	(0.000672)	(0.000428)	(0.000729)
Age 55-64	-0.0534***	-0.0487***	-0.00201***	-0.00668***
	(0.000956)	(0.000775)	(0.000436)	(0.000691)
Married	-0.0202***	-0.0132***	0.00724***	0.00757***
	(0.000652)	(0.000473)	(0.000280)	(0.000553)
Widowed	-0.0361***	-0.0124***	0.00513***	0.0141***
	(0.00134)	(0.000969)	(0.000658)	(0.00112)
Divorce/Separate/Annulled	-0.0214***	-0.00430***	0.00703***	0.0278***
	(0.00177)	(0.00112)	(0.000889)	(0.00196)
Observations	1437827	2553367	2448570	2496782
R-squared	0.0536	0.0162	0.0383	0.0428
Joint Significance (F-stat)	333.57	178.40	173.48	144.22
Region FE	Yes	Yes	Yes	Yes
Year & Quarter FE	Yes	Yes	Yes	Yes

Notes:

(1) All standard errors (in parentheses) are heteroskedasticity robust;

(2) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$;

(3) Joint significance (F-stat) measures the joint significance of the regressors, including the identified factors, their interactions with region dummies, and demographic dummies.

$EN_{Nonagri}$ and NE_{Agri}. Specifically, for a 1% increase in real wage differentials, on average, the probability of flowing into non-employment from agriculture increases by approximately 0.01 percentage point, whereas the probability of flowing out of non-employment into agriculture decreases by 0.007 percentage point. This implies that widening real wage differentials give rise to non-negligible incentives for workers to switch out of agriculture and direct their search effort elsewhere. Moving on to labor market efficiency and road density, we find that both of them have a positive and highly significant effect on job separations out of agriculture and job findings into non-agriculture. This implies that the industry switching motive becomes especially prominent when labor market efficiency is improving in non-agriculture and transport infrastructure is upgraded, as agricultural workers can switch industries more freely.

In terms of high school and college education, we find that both of them contribute to employment outflows from agriculture through higher EN_{Agri} and lower NE_{Agri}; however, in levels, college education plays a more significant role in driving the outflows than high school education. This confirms our proposition in Figure 2.10 that the lack of sufficient education, especially at higher levels, suppresses employment outflows from agriculture and essentially drags workers back to agricultural employment. Furthermore, if we compare the AMEs of college education on $EN_{Nonagri}$ and $NE_{Nonagri}$ to those of high school education, we find that it is only by undertaking college education that workers can find and hold on to their non-agricultural jobs more easily.[11] Hence, given that only less than 10% of employed workers in agriculture reach college education, this further attenuates employment outflows from agriculture to non-agriculture.

As for the regional share of total palay and corn production, its negative and significant AMEs on EN_{Agri} and $NE_{Nonagri}$ unequivocally indicate the negative impacts of agricultural clusters, as the difficulty to switch out of the agricultural

[11]This seems to be driven mostly by transition rates for high-skilled non-agricultural jobs. For low-skilled non-agricultural jobs, having college education, as in the case of agriculture, incentivizes workers to separate more and find less. Nonetheless, high school education seems to have a larger impact on both job separation and finding rates in agriculture than those in low-skilled non-agriculture. See the third robustness check in Section 2.3.3 for details.

sector is heightened when the region's economy is more dominated by agriculture. However, the regional share of total palay and corn production does not seem to have a significant impact on the outflows of non-agricultural employment or inflows into agricultural employment.

Lastly, by examining the AMEs of demographic dummies, we find that the gender bias is present in the Philippines as females separate significantly more often and are less likely to find jobs in both agricultural and non-agricultural sectors. Furthermore, though not uniformly, workers become significantly less likely to separate when they age. The latter observation is likely to be associated with experience and learning about industrial fit as in, for instance, Gervais et al. (2016) and Menzio et al. (2016). The intuition is that, in order to learn the industry in which they are most productive and earn the most wages, workers sample industries over their careers. Because young workers are more likely to be in matches that represent a poor industrial fit, they spend more time in transitioning between industries. As for old workers, since they have drawn more matches, they are more likely to be in ones with higher idiosyncratic productivity, and hence, less likely to separate. In this way, we expect a declining age profile of job separation.

2.3.3 Robustness Checks

In the previous section, we identify several 'push' and 'pull' factors in explaining the observed changes in the EN and NE rates. In what follows, we elaborate on the empirical results along several dimensions. We aim to apply the following exercises as robustness checks and to further illustrate the driving forces of the agricultural exodus in the Philippines.

2.3.3.1 Cross-Region Wage Differentials and Industry Switching via Migration

Geographic mobility is a potentially powerful way of escaping non-employment and improving one's income. The literature (e.g., Kennan and Walker (2011)) have indicated that migration decisions are influenced to a substantial extent by income prospects. The link between income and migration decisions is driven both

by geographic differences in mean wages and by a tendency to move in search of a better locational match when the income realization in the current location is unfavorable. This finding leads us to speculate how much of the intersectoral labor reallocations in the Philippines are accompanied by geographic reallocations.

Although the LFS data do not allow us to track each individual beyond two periods and across regions, we estimate how much the effects of cross-region real wage differentials are on the labor market transition rates, and show indirectly that some job separations/findings are indeed triggered by them, as individuals migrate from one region to another in order to switch industries. Specifically, we calculate both the within-region and cross-region real wage differentials. The within-region real wage differentials relate to the same measure as the one applied in the main regressions, which is the natural logarithm of the differences in weighted average real hourly wages between non-agricultural and agricultural workers in each region, i.e., $w_{x,Nonagri} - w_{x,Agri}$ for region x. In terms of cross-region real wage differentials, the definition varies depending on whether the dependent variable (i.e., EN or NE rate) is for agricultural or non-agricultural workers. More specifically, when we study the transition rates for agricultural workers, the cross-region real wage differentials in region x can be calculated as the natural logarithm of the differences between weighted average real hourly wages earned by non-agricultural workers in all other regions and those earned by agricultural workers in region x, i.e., $w_{\neg x,Nonagri} - w_{x,Agri}$. For non-agricultural workers, it is computed as $w_{x,Nonagri} - w_{\neg x,Agri}$ for region x.[12] In this way, we are able to capture the magnitude of geographic wage differentials, and study the responsiveness of intersectoral labor flows to them.

Table 2.2 gives the estimation results. We see that after taking into account cross-region real wage differentials, the AMEs of the original 'push' and 'pull' factors, including within-region real wage differentials, are robust in general. As for cross-region real wage differentials, their marginal effects are highly significant, and the signs imply that greater cross-region real wage differentials give rise to more

[12]Such calculations make the AMEs comparable across transition rates in agriculture and non-agriculture, as any increase in the wage differentials implies a widening wage gap between non-agricultural and agricultural workers.

job separations and findings in the agricultural sector, whilst fewer separations but more findings in the non-agricultural sectors. In particular, the positive effect of cross-region real wage differentials on NE_{Agri} points to the likely circumstance that significant employment outflows from agriculture lessen the competition within, and thus result in higher job finding rate in agriculture. In addition, thin cross-regional non-agricultural labor markets, such that the actual cost of migration outweighs the potential benefit, can also reduce geographic mobility and essentially make agricultural workers move less (Balgova, 2020).

Table 2.2: AMEs in the LPM with cross-region real wage differentials

	Agriculture		Non-Agriculture	
	EN_{Agri}	NE_{Agri}	$EN_{Nonagri}$	$NE_{Nonagri}$
Within-Region Real Wage Differentials	0.0112***	-0.0147***	-0.00334	0.0108***
	(0.00305)	(0.00132)	(0.00247)	(0.00241)
Cross-Region Real Wage Differentials	0.0106*	0.0328***	-0.00945***	0.0232***
	(0.00546)	(0.00222)	(0.00362)	(0.00348)
Labor Market Efficiency	0.0122***	0.00135	-0.0121***	0.0416***
	(0.00293)	(0.00172)	(0.00298)	(0.00196)
Road Density	0.0198	-0.0956***	-0.0283	0.0266
	(0.0225)	(0.0120)	(0.0203)	(0.0186)
High School Education	0.0204***	-0.0195***	0.00275***	-0.00238***
	(0.000473)	(0.000352)	(0.000469)	(0.000476)
College Education	0.0320***	-0.0275***	-0.0118***	0.00174***
	(0.000935)	(0.000342)	(0.000482)	(0.000511)
Palay & Corn Production	-0.104***	0.00536	-0.0222	-0.0370**
	(0.0172)	(0.00903)	(0.0171)	(0.0155)
Observations	1437160	2380033	2551595	2491172

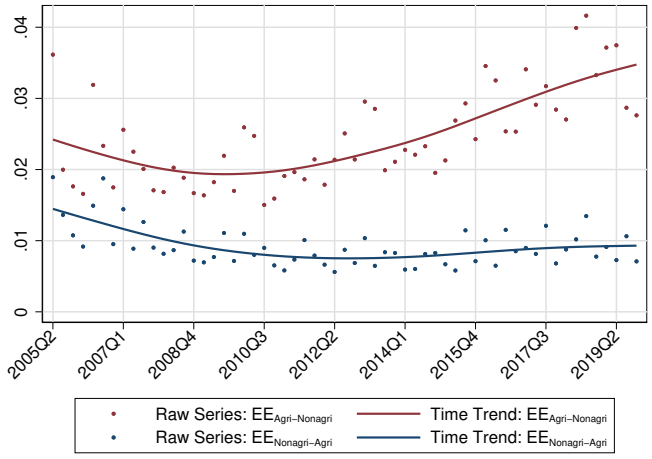
Notes:
(1) All standard errors (in parentheses) are heteroskedasticity robust;
(2) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$;
(3) Demographic, region, and year & quarter dummies are included;
(4) The definitions of cross-region real wage differentials differ between agricultural and non-agricultural workers; see the above text for details.

2.3.3.2 EE Transitions

Despite the rare occurrence of EE transitions in the LFS, they still represent an integral part of labor market flows in theory (see Figure 2.2). As presented in Figure 2.11, by inspecting time series trends of the EE rates, we find that in

levels, the EE rate from agriculture to non-agriculture is higher than the opposite; both trends are decreasing in the early sample before they diverge as the EE rate from agriculture to non-agriculture rises significantly, while the EE rate from non-agriculture to agriculture remains relatively stable.

Figure 2.11: Time series trends of the EE rates



Source: Labor Force Survey, Philippine Statistics Authority. The trends are estimated using the Hodrick-Prescott Filter with smoothing parameter of 1600.

We then estimate how much the proposed 'push' and 'pull' factors explain the observed evolution in the EE rates. The results are summarized in Table 2.3. We firstly find that only the AME of real wage differentials on EE transitions from non-agriculture to agriculture is negative and statistically significant. This implies that as wage differentials widen, non-agricultural workers become more reluctant to separate, no matter whether it is separation into non-employment (as also captured by the negative and statistically significant AME of real wage differentials on $EN_{Nonagri}$ in Table 2.1), or separation to agricultural employment. In contrast, real wage differentials do not seem to affect the job-to-job transitions out of agriculture. Combined with that the marginal effects of labor market efficiency, high school, and college education are positive and statistically significant, this result highlights the importance of providing means to agricultural workers so as to enhance their ability to switch industries freely.

Table 2.3: AMEs in the LPM for EE transitions

	$EE_{Agri \to Nonagri}$	$EE_{Nonagri \to Agri}$
Real Wage Differentials	-0.00168	-0.00298***
	(0.00183)	(0.000750)
Labor Market Efficiency	0.0121***	-0.00872***
	(0.00164)	(0.00184)
Road Density	-0.00211	-0.0132
	(0.0142)	(0.00845)
High School Education	0.00749***	-0.0103***
	(0.000340)	(0.000238)
College Education	0.00821***	-0.0157***
	(0.000629)	(0.000225)
Palay & Corn Production	-0.0314**	-0.0283***
	(0.0129)	(0.00630)
Observations	1386682	2450249

Notes:
(1) All standard errors (in parentheses) are heteroskedasticity robust;
(2) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$;
(3) Demographic, region, and year & quarter dummies are included.

We also find through the negative AMEs of the regional share of total palay and corn production that workers cling to their jobs, no matter whether they are agricultural or non-agricultural, when the regional economy is more dominated by agriculture. This is expected as on the one hand, agricultural workers face more difficulties to switch out of agriculture when the cluster effect takes hold; on the other hand, non-agricultural workers become more reluctant to separate due to the wages loss from job separation and limited re-employment opportunities back into non-agriculture.

2.3.3.3 Switching From Agriculture to Other Low-Skilled Industries

In this section, we study the transition rates in low-skilled non-agricultural industries instead of the non-agricultural sectors as a whole. This exercise adds to the discussion that the reallocations of agricultural workers into low-skilled non-agricultural industries may be more practically feasible. First of all, there might be large static productivity gains when labor reallocates from agriculture to non-agricultural

industries. Sizable productivity and wage gaps between agriculture and non-agricultural industries have been measured in several studies and have been shown to be larger in developing economies (e.g., Gollin et al. (2014), Lagakos and Waugh (2013), Restuccia et al. (2008), and Caselli (2005)). To the extent that these gaps arise from the existence of inefficiencies and frictions in the economy, reallocations of labor from agriculture to non-agricultural industries are both productivity- and welfare-enhancing.[13] Secondly, there can be dynamic productivity gains when labor reallocates towards non-agriculture if this industry is subject to agglomeration externalities and knowledge spillovers (e.g., Matsuyama (1992)). Lastly, inducing labor reallocations from agriculture towards low-skilled non-agricultural industries can be more practically sound than reallocations to high-skilled industries, as the skill requirements might be less demanding for such transitions to take place.

We define low-skilled non-agricultural industries as those which have an average share of employed workers with at least high school education below 70%, i.e., mining and quarrying, construction, and domestic and household services. We then estimate the average marginal effects of the proposed 'push' and 'pull' factors on the transition rates in agriculture and low-skilled non-agricultural industries respectively. The results are summarized in Table 2.4. We find that changes in labor market efficiency are driving the employment outflows from agriculture towards those low-skilled industries, whilst real wage differentials and road density only play a minor role. As for high school and college education, they accelerate employment outflows from the agricultural sector mainly through increasing the separation margin and shrinking the job finding probability in agriculture. Here, an interesting observation is that while the job separation rate in low-skilled non-agricultural industries increases with high school and college education, the job finding rate in these industries also decreases with them. The latter differs from the job finding rate in non-agriculture as a whole (see Table 2.1), reflecting the role of college education in high-skilled industries. As for its cause, we argue that it is directly

[13] Although this view has been recently challenged by Bustos et al. (2020) that low-skilled workers leaving agriculture reinforces comparative advantage in the least skill-intensive manufacturing industries, which reduces the incentives to innovate and slows down growth.

related to the strong incentive for skilled labor to switch away from low-skilled industries in general. Nonetheless, there are large differences in the size of AMEs of high school and college education on labor market transition rates between agricultural and low-skilled non-agricultural sectors.

Table 2.4: AMEs in the LPM for transitions from agriculture to other low-skilled industries

	EN_{Agri}	$EN_{Lowskilled}$	NE_{Agri}	$NE_{Lowskilled}$
Real Wage Differentials	-0.00231	-0.00515	-0.00334***	-0.00272***
	(0.00196)	(0.00473)	(0.000750)	(0.000611)
Labor Market Efficiency	0.0115***	-0.0351**	-0.00193	0.0104***
	(0.00302)	(0.0165)	(0.00161)	(0.000887)
Road Density	-0.00373	0.146**	-0.0240**	0.00528
	(0.0222)	(0.0567)	(0.0106)	(0.00788)
High School Education	0.0204***	0.00357***	-0.0191***	-0.00534***
	(0.000473)	(0.00113)	(0.000345)	(0.000277)
College Education	0.0320***	0.00621***	-0.0269***	-0.0100***
	(0.000935)	(0.00167)	(0.000335)	(0.000269)
Palay & Corn Production	-0.106***	-0.123**	0.0190**	0.00159
	(0.0162)	(0.0484)	(0.00809)	(0.00727)
Observations	1437666	394678	2430806	2407870

Notes:
(1) All standard errors (in parentheses) are heteroskedasticity robust;
(2) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$;
(3) Demographic, region, and year & quarter dummies are included.

2.3.3.4 Other Robustness Checks

We also perform three other robustness checks with different regression specifications. Firstly, we explore the extent to which our baseline regression results are affected during specific time periods. We apply two exercises in Appendix A.3, in which we add a time dummy indicating 2005Q2-2008Q1 and 2017Q1-2019Q4 respectively in the LPM of each labor market transition rate, and interact the time dummy with the proposed 'push' and 'pull' factors. We find that the results are robust in general, with no significant differences in the coefficients of 'push' and 'pull' factors over time. Secondly, we report in Table A.3 the estimation results corresponding to a specification with clustered standard errors at the regional level. We find that

even though real wage differentials and road density do not significantly explain the variations in the EN and NE rates for either agricultural or non-agricultural workers anymore, the remaining 'push' and 'pull' factors still have their explanatory power, especially for high school and college education. Lastly, we re-estimate the AMEs of each of the proposed 'push' and 'pull' factors using a logit model. The results are largely similar to what we observe in the LPM, and are presented in Table A.4.

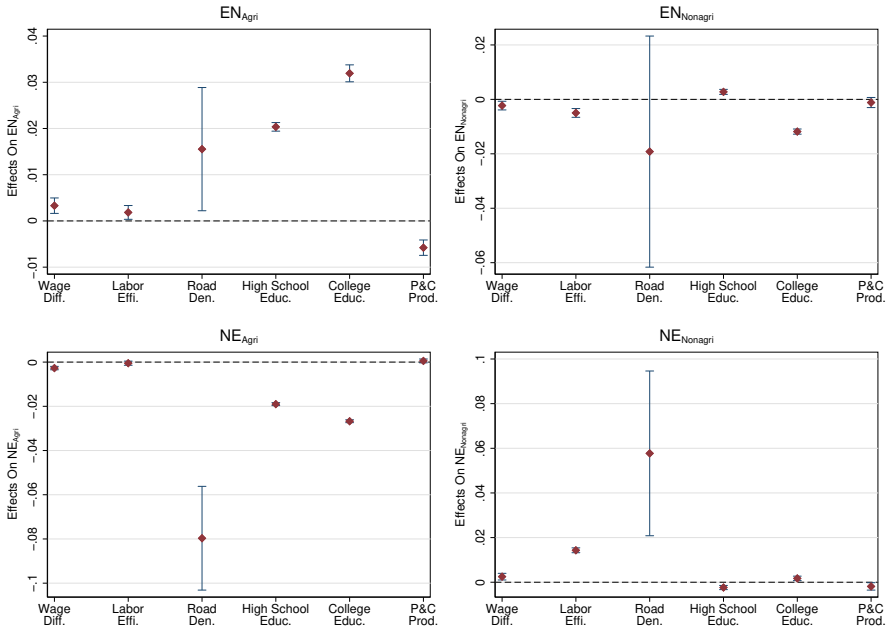
2.4 Which Drivers Have the Largest Impact?

As explained in the previous section, each of the proposed 'push' and 'pull' factors is statistically significant in explaining the observed variations in the labor market transition rates. However, which of them are more economically significant in driving the employment outflows from agriculture to non-agriculture? To answer this question, we calculate the economic significance of each factor as to how much the EN or NE rate will change per standard deviation increase in the identified factor. For high school and college education, since they are dummy variables, we measure their economic significance by the corresponding AMEs that are presented in Table 2.1. The results are shown in Figure 2.12. From the panels, we see that real wage differentials are highly statistically significant but the impacts are very limited.[14] In contrast, high school and college education are more critical, especially for job separation from agriculture. Similarly, road density plays a more economically significant role, especially for job finding in both agriculture and non-agriculture.

The above exercise emphasizes the importance of promoting post-primary education in order to facilitate intersectoral labor reallocations. For agricultural workers, returns to off-farm labor are constrained by their low levels of education and by the large disparities in higher education attainment between agricultural and non-agricultural workers (recall Figure 2.10). Since the main dividing line between high- and low-paying jobs is skills, more educated workers are more likely to have access to rewarding non-agricultural jobs. In addition, considering the regional differences in

[14]We also calculate the economic significance of cross-region real wage differentials, and show that it is similar to that of within-region real wage differentials.

Figure 2.12: Economic significance



Source: Labor Force Survey, Philippine Statistics Authority.

higher education attainment, from Figures A.4, we see that the effect of high school education in driving employment outflows from agriculture is general across regions, but often even larger for regions with less education through higher EN_{Agri} and lower NE_{Agri}. This implies that public investment in education can lean towards those underdeveloped regions with low access to high school education. In addition, we find in Figure A.5 that promoting high school education, especially among the young, leads to more significant labor reallocations through higher job separation and lower job finding in agriculture.[15] Furthermore, females benefit proportionally more from higher education as they experience both significantly higher job separation in agriculture and job finding in non-agriculture than males when undertaking high school education (see Figure A.6). Lastly, we take a step forward by examining the effects of college education on intersectoral labor reallocations. By comparing the AMEs of college education on labor market transition rates across regions, age, and gender in Figures A.7, A.8, and A.9 respectively, we find similar results to those

[15]However, the effect could be limited as the problem of aging population in agriculture takes hold (i.e., workers with age from 35 to 64 represent a significant proportion of labor employed in the agricultural sector).

associated with only high school education in terms of transitions in and out of agriculture; yet, as for transitions in and out of non-agriculture, it is worth noting that college education seems more effective for workers (especially the females) to find and keep their non-agricultural jobs.

In addition to higher education, improving transport infrastructure also turns out to be a practical and economically significant option of facilitating labor reallocations from agriculture. We show in Figure A.10 that, though not uniformly, in regions with low levels of road density, the marginal increase in road density induces larger outflows from agriculture to non-agriculture (higher EN_{Agri} and $NE_{Nonagri}$, in general). This result is consistent with the broad literature arguing that transport infrastructure promotes intersectoral reallocations, and thus increases welfare, especially in rural and underdeveloped areas (e.g., Donaldson (2018) and Y. Yang (2018)). It is also in line with the growing interest among policymakers worldwide in using infrastructure investments both as a short-term fiscal policy instrument, and more importantly, as a long-term growth generator, which calls for better resource reallocations between industries with significant productivity differences.

2.5 Conclusion

In this paper, we set our path to examine the underlying reasons for the agricultural exodus that took place in the Philippines during the last 15 years. In the estimation of labor market transition rates using the Philippines LFS, we firstly illustrate that the shrinking agricultural employment can be summarized as a flow process transitioning from agriculture to non-agriculture; we show empirically that workers separate more from their agricultural jobs and increasingly direct their search effort to non-agriculture. Next, by applying both the theoretical and empirical frameworks, we find evidence that some 'push' factors, namely, widening wage differentials, increasing labor market efficiency, and expanding road density in the Philippines are significantly correlated with employment outflows from agriculture, whilst the 'pull' factors, i.e., lack of high school and college education attainment and agricultural clusters, are associated with dampening the process. Furthermore,

we estimate the economic significance of each proposed factor and find that in contrast to the traditional view, increasing wage differentials do not explain much of the agricultural exodus; instead, measures related to enhancing the worker's ability to freely switch industries exert more influence on their industry switching decision. In particular, we show that even high school education can add significantly more to the exodus through raising agricultural worker's job separation margin substantially. In addition to that, when college education is undertaken, employment outflows from agriculture to non-agriculture could be further amplified because workers can find and hold on to non-agricultural jobs more easily. Improving transport also plays a more economically significant role in driving intersectoral labor reallocations, as not only does higher road density reduce job finding in agriculture, but also it boosts job finding in non-agriculture.

As previously indicated, our empirical study only selects a non-exhaustive list of factors based on the discipline provided by the model and estimates their correlation with the observed evolution in the labor market transition rates. Although our selection is restricted by the availability of data, we believe that the augmented theoretical framework could offer a more comprehensive view of what the 'push' and 'pull' factors are in general, and more importantly, through what mechanism these factors give rise to increasing employment outflows from agriculture to non-agriculture. In addition, the empirical framework could be easily extended to account for other influential factors once their data become available. Two notable examples include the land reform progress in the Philippines (Adamopoulos & Restuccia, 2020) and climate change (Rosegrant et al., 2016), both of which have the potential to affect agricultural productivity and hence the pace of employment outflows from agriculture. We leave these extensions for future endeavors.

3

A Big Data Approach to Cargo Type Prediction and Its Implications for Oil Trade Estimation

3.1 Introduction

Economic booms and recessions are often accompanied by trade fluctuations (Cerdeiro et al., 2020; Kohn et al., 2021; Mendoza, 1991). The traditional trade data that are required to assess such fluctuations, however, are often only available from international organizations or through the customs bureaus of individual countries. These data are mostly low frequency (published monthly or yearly), delayed (with two to three months of lag), and lacking adequate public access. Given such limitations, there has been a growing interest in maritime literature towards constructing high-frequency trade data that provides real-time information on global trade activity (Arslanalp et al., 2019; Cerdeiro et al., 2020). In theory, such high-frequency trade data could capture rapid changes in the world economy, which would significantly improve the ability of policymakers to identify emerging trade flow risks and detect potential turning points in the business cycle.[1] In addition, these data are also instrumental for corporations entwined in the global

[1] High frequency trade data are of vital importance for policymaking, especially during significant economic downturns (e.g., the ongoing COVID-19 pandemic).

supply chain system, since they enable those corporations to assess risks promptly, make timely decisions, and better manage or control the supply process.[2]

Transportation data are a natural candidate for a guide to a comprehensive picture of the global economy. Among all transportation modes, maritime shipping represents an integral part of global trade: It constitutes over 80 percent of global trade by volume, and more than 70 percent by value (United Nations Conference on Trade and Development, 2019). Therefore, by digging into seaborne transportation and examining its working mechanisms, we can study the dynamics of trade activity and quantify them accordingly. In this paper, we resort to the Automatic Identification System (AIS) data on coated product tankers, as this information provides unprecedented details of these tankers' real-time movements across the globe. This makes it possible to observe retrospectively the behaviors of each tanker, identify their loaded trips, and estimate the corresponding oil trade volume at a high frequency.

The attempt to apply AIS data on oil tankers to the problem of crude oil trade estimation is not new in the maritime literature. Nonetheless, there has been almost no attempt to study the market switching behavior of coated product tankers in a systematic manner, or to consider its significant implications for global crude oil trade estimation. Specifically, in the oil transportation industry, in addition to crude oil tankers (which are vessels dedicated to serving the crude oil market), coated product tankers (which are designed to carry clean petroleum products (CPP), including naphtha, gasoline, gas oil, and jet fuel), can also opportunistically switch to the trade in dirty petroleum products (DPP), the majority of which are crude

[2]For instance, real-time sales and trade data could enable multinational retailers and manufacturers to make more informed decisions regarding whether to regionalize their supply chains in response to supply bottlenecks, soaring transportation costs, and other disruptions caused by the COVID-19 pandemic. See https://www.ft.com/content/6c127633-9825-4921-9a11-71518591e2f3 (Accessed 13 September 2021) for a more detailed discussion of this issue.

oil.[3,4,5] This switching behavior takes place very often in practice, as seen in the fixture data provided by Clarksons Shipping Intelligence Network (which only record a limited fraction of crude oil trips taken by coated product tankers): Among the 4,748 matched fixtures recorded from 2017-2020, more than one-third (1,568) carried crude oil, while the rest carried either CPP or dirty refined product oil. Nevertheless, despite the regularity of crude oil trips made by these product tankers, existing studies either consider only crude oil tankers in their estimation of global seaborne crude oil trade (Cerdeiro et al., 2020), or simply apply external information (such as oil import and export attributes of a port (Yan et al., 2020) or line-up reports (Adland, Jia, et al., 2017)) in order to distinguish between crude and refined product oil trips taken by product tankers.[6] Therefore, in order to quantify the global crude oil trade more accurately and without heavy reliance on external information, it is crucial to develop a structured and easily applicable methodology that can classify the cargo type of a coated product tanker during each of its identifiable loaded trips.

To achieve this goal, we apply the random forest (RF) ensemble learning technique (Breiman, 2001) to the AIS data of 752 coated product tankers in order to classify their cargo types. Specifically, we first identify the loaded trips taken by each product tanker during the period from 2017-2020. Subsequently, based on

[3]It should be noted that such a switching behavior involves a one-off, tank cleaning cost to the shipowner when it re-enters the CPP trade. This is because, by industrial practice, the ship's cargo tank needs to be cleaned before it loads CPP in order to avoid the risk of cargo admixture contamination.

[4]Due to the corrosive nature of many refined petroleum products, special-purpose product tankers are designed with coated cargo tanks, meaning that they can be used to carry either crude or refined product oil. In this paper, we focus on the coated Long Range (LR) class product tankers. Specifically, LR1 ships are long range product tankers with a capacity of between 55,000 and 84,999 deadweight tonnage (DWT), whereas LR2 ships are those with a capacity of between 85,000 and 124,999 DWT.

[5]On the one hand, Adland, Hansson, et al. (2017) attribute the market switching behavior of product tankers to their profit-driven motive in response to the freight rate differentials between the product and crude oil trades. Similarly, Bai and Li (2021) propose a search and matching framework that encompasses the market switching behavior of product tankers while being unloaded ("ballasting"); their setup highlights that unloaded ships ballast in response to differences in the market-specific payoffs. On the other hand, Poten and Partners (2014) argues that technical and organizational factors can also affect the decision to switch, e.g., whether the underlying vessel has new or well-maintained tank coatings.

[6]Information on cargo type is usually contained in line-up reports and shipping contracts, both of which are often only available to the parties involved in the transaction but not to the general public. This information is not contained in the AIS data either.

domain knowledge obtained from extensive consultation with industry experts, we construct a set of input variables tailored to our prediction problem. Next, we use the RF model to predict the cargo type of a product tanker during each of its loaded trips identified in the AIS. This involves two stages: We first apply the RF model to distinguish between CPP and DPP trips made by the product tankers during the sampling period; then, for those estimated DPP trips, we re-apply the model to further distinguish between crude and dirty refined product oil trips. When its predictive ability is tested against that of alternative models, the RF model performs the best in several evaluation metrics.

Based on the resulting shipping dataset, we establish that, among the 23,729 loaded trips identified in the AIS from 2017-2020, 5,923 (25.0%) can be classified as crude oil trips, 5,130 (21.6%) as dirty refined product oil trips, and the remaining 12,676 (53.4%) as CPP trips. More importantly, the 5,923 crude oil trips represent approximately 8% of the global seaborne crude oil trade volume during 2017-2020, indicating the indispensable role of coated product tankers in the transportation of crude oil worldwide.[7] Furthermore, as an external validation of our estimation strategy, we perform a quantitative exercise that studies unanticipated variations in the crude oil volume carried by product tankers. We show that such variations are consistent with several major oil trade disruptions that took place between 2017 and 2020.

The main contributions of our paper are threefold. First, we develop an RF-based prediction model of cargo types for product tankers based on the AIS and other externally available data. To the best of our knowledge, our model is the first in the literature that explicates the market switching behavior of coated product tankers, and applies an ensemble learning approach to estimate their cargo types.[8] Second, the rational selection and combination of input variables are another contribution of the big data approach developed in this paper; they add to the existing domain

[7]Data on the global seaborne crude oil trade volume from 2017 to 2020 are extracted from Statista (2021).

[8]Our application of the RF ensemble learning technique to cargo type prediction is itself another contribution to the literature, since it is accepted (e.g., Chen et al. (2017)) that when weak models are combined correctly, more accurate and/or robust prediction models can be attained.

knowledge within the field of maritime transportation, deepen our understanding of the mechanics of oil shipping markets, and facilitate the granular exploration of the characteristics of oil transportation. Lastly, the shipping dataset generated from our predictions enables the quantification of the global crude oil trade in a more accurate and tractable manner.

Our model and the associated decomposition strategy could be utilized by different stakeholders in various applications. First, policymakers could easily apply them to identify emerging crude oil trade disruptions and craft sound policies, e.g., preemptive release of oil reserves in anticipation of soaring oil price due to an oil trade disruption. This would not be possible should we neglect the endogenous choice by the shipowner and instead assume the LR tanker always carries crude or product oil, as the change in tanker utilization (i.e., the proportion of tankers that are loaded with oil cargoes in a shipping market) in the crude oil market that is induced by a crude oil shock might be camouflaged with a corresponding change in tanker utilization in the product oil market when shipowners make a switching decision, hence leaving the oil trade volume unchanged. The proposed model could also help promote data transparency in the maritime industry because the type of product a vessel carries used to be blackbox information and not known to the industry. Although there are specialized data providers collecting such information, it is often very costly to acquire and only covers a limited fraction of fixtures in the market. Without loss of generality, the proposed modeling framework could be extended to different vessel types, and it is expected to greatly enhance the shipping data transparency at a global scale with relatively low costs. Lastly, the results have significant implications for major players in the shipping market, including charterers, shipowners, and traders. For charterers, they could utilize the model to analyze how tankers operated by different shipowners change their market switching behaviors under various market conditions, and thus be able to make more informed chartering decisions in advance. For shipowners, the model could be applied to analyze their competitors' operational strategies, and to benchmark their own performance accordingly. They could also make better decisions on which market

to enter for the next contract by estimating the proportion of vessels engaging in each type of product trades. As for traders, they could use the model to estimate the (near) real-time, global, aggregate trade volume of different product oil, and hence develop a better understanding of the real-time dynamics in oil trade.

The rest of this paper is organized as follows. Section 3.2 provides a review of the literature on the AIS data and their applications in academic research, with particular reference to current studies that apply the AIS data of oil tankers to estimate the global crude oil trade. Section 3.3 describes the sources of the data used in this paper. Section 3.4 presents the data processing techniques that are applied to identify loaded trips. Section 3.5 lays out the details of the RF model in the estimation of cargo types. Section 3.6 illustrates our prediction results, and how they compare to those of alternate models. Section 3.7 presents an application of the resulting shipping dataset to a study of unanticipated fluctuations in the volume of crude oil transported by product tankers from 2017-2020. Section 3.8 concludes. The Appendices contain additional tables, figures, and information on alternative prediction models.

3.2 Literature Review

The AIS data have been at the center of much research in maritime transportation. Originally designed to avoid ship collision, the AIS is an automatic tracking system installed on ocean-going vessels. Its records provide real-time satellite and terrestrial data on a ship's static (e.g., identity and size), dynamic (e.g., position, speed, and time stamp), and voyage-related (e.g., draught, destination, and estimated time of arrival (ETA)) features (D. Yang et al., 2019). They have been applied in various research areas, including but not limited to, ship collision and navigation safety (W. Zhang et al., 2015), shipping behavior analysis (Breithaupt et al., 2017; Kroodsma et al., 2018), environmental evaluation and ship emission control (Bai et al., 2021; Simonsen et al., 2019; Winther et al., 2014), shipping supply and freight rate prediction (Regli & Nomikos, 2019), and commodity trade estimation

(Adland, Jia, et al., 2017; Cerdeiro et al., 2020; Jia et al., 2019; Prochazka et al., 2019; Yan et al., 2020).

Recently, AIS-driven trade analysis has received increasing attention from the academy. In particular, several existing studies focus on estimating oil trade flows from the AIS data of oil tankers.[9] As argued earlier, however, despite the informativeness of AIS data in the estimation of global crude oil trade volume, current research tends either to consider only crude oil tankers or to simply apply externally available information in order to distinguish between crude and refined product oil cargoes transported by product tankers. For instance, Cerdeiro et al. (2020) propose a real-time trade flow monitoring method based only on the AIS data of crude oil tankers, and estimate that the correlation between the volume of crude oil imports calculated using the AIS data and that of the Joint Organizations Data Initiative (JODI) statistic is 0.73 in the raw data, and 0.81 on a 3-month moving average basis. The study by Yan et al. (2020) shows similar results: By developing a ship cargo payload calculation model, they also find that the oil trade volume of the world's top 20 oil-importing and oil-exporting countries calculated based on the AIS data is strongly correlated with the JODI statistic, with a coefficient of determination (R-squared) at 0.8798.[10] In the model, they apply information regarding the oil import and export attributes of a port to distinguish between oil cargo types. Specifically, they first classify a port as oil-importing if the total amount of oil cargoes unloaded by oil tankers entering the port exceeds that of cargoes loaded there, and oil-exporting if otherwise. Given the port designations, they then classify the cargo carried by an oil tanker as refined product oil if it is loaded at an oil-importing port. Otherwise, the oil tanker is deemed to be carrying crude oil. Such a strategy of classifying cargo types has the advantage of simplicity, although

[9] To compute trade flows, two inputs from the AIS are normally required: (1) a vessel's payload, i.e., cargo quantity carried by a vessel (which can be estimated using its draught information), and (2) its spatial displacements. In our method, these two inputs form the basis for the identification of loaded trips.

[10] In addition, Yan et al. (2020) illustrate that on average, the oil trade volume calculated using the AIS data is 24% lower than the JODI statistic. They attribute this finding to the fact that the pipelines, which are the major alternative transportation mode for crude oil, are not considered in the AIS data.

its accuracy suffers because, in reality, not all cargoes loaded at an oil-importing port are refined product oil. Finally, Adland, Jia, et al. (2017) also take advantage of ship positioning data from the AIS in order to derive more timely and detailed crude oil trade statistics. Specifically, they show a detailed breakdown of the crude oil trade by vessel size over time, and find that the AIS-derived data for seaborne crude exports show a good alignment with official export numbers in aggregate – a conclusion similar to Yan et al. (2020). As indicated in their paper, however, their shipping data result from a combination of the AIS-tracking of individual vessel voyages with cargo information from the line-up reports produced by port agents, which are often only available to the parties involved in the transaction and not to the general public. In sum, these limitations of the existing studies render their estimations of the crude oil trade through AIS data either less accurate or less replicable. Our paper directly addresses these two angles by developing a tractable algorithm on the basis of supervised machine learning that demonstrates a strong capacity to predict the cargo types of product tankers.

In addition to trade analysis, there is another strand of AIS-driven literature that studies vessels' micro behaviors, even though their research focuses are distinct from ours. In particular, C. Zhang et al. (2020) propose an RF-based approach that applies the AIS data to predicting a vessel's destination. In their model, the similarities between a vessel's currently traveling and historical trajectories are measured, and the destination along a historical trajectory that shows the greatest similarity to the traveling trajectory, is predicted as the vessel's destination. Another example is Alessandrini et al. (2018), who take advantage of the AIS data in order to develop an optimized, data-driven, path-finding algorithm that estimates a vessel's ETA. In a similar vein, Sugrue and Adriaens (2021) propose a maritime transport efficiency (MTE) metric by integrating the AIS information with navigation lock data, and presenting a mathematical model to predict a vessel's payload from variations in water surface elevations. In addition, the pioneering work of Brancaccio et al. (2020) applies the AIS data to a study of the interactions between dry bulk ships and exporters. They present novel evidence of search frictions in the

dry bulk markets, and offer unique insights into the role of the transportation sector in attenuating differences in comparative advantages across countries, creating network effects in trade costs, and dampening the impacts of shocks on trade flows. In this paper, we add to the existing literature by studying the market switching behavior of product tankers at the micro level.[11] In addition to its application to crude oil trade estimation, Bai and Li (2021) also employ the shipping data derived from AIS to study how the micro behaviors of oil tankers and their interactions with exporters determine congestion in the shipping markets, an understanding that is critical both for the international oil trade and the macroeconomy. Similarly, our paper explores the ramifications of the micro behaviors described in our model for the prediction of macroeconomic trends useful for stakeholders across the globe.

3.3 Data

In order to construct the shipping dataset that underlies our prediction of cargo types for product tankers, we consider two main data sources. The primary source is the AIS data of 752 coated product tankers, which cover the entire coated product tanker fleet during the period of 2017-2020. In addition, we apply fixture data from Clarksons Shipping Intelligence Network collected during the same sampling period. The data record a non-exhaustive list of transactions between shipowners and exporters and their respective features, such as vessel name, laycan (laydays canceling, or L/C) dates, loading and discharge ports, shipowner, cargo type, etc. Fixture data are essential in our study, as the AIS data do not provide information on the type of oil cargo a ship carries in a loaded trip.

Nevertheless, our fixture data only record the broker's own fixtures (i.e., deals closed by Clarksons) and hearsay in the market. Therefore, such data only capture a

[11] There has only been limited research on vessels' market switching behavior, not to mention the associated trade implications. Notably, Sødal et al. (2008) is one of the few papers that studies the market switching behavior of vessels, and in which they derive a real options model of flexibility, apply it to shipping, and value the option of switching between the dry and wet bulk markets for a combination carrier. Furthermore, Adland, Hansson, et al. (2017) apply an entry-exit real option model with a stochastic freight rate differential in order to derive the optimal triggers for switching between the two cargo types for product tankers; they estimate the value of the switching option accordingly.

limited fraction of the transactions that take place across the entire shipping market of product tankers. For those remaining loaded trips with an unclassified cargo type, we apply the RF ensemble learning technique so as to estimate their cargo types during each of the loaded trips. The estimation details are elaborated in Section 3.5.

3.4 Identification of Loaded Trips

Having described the data sources, in this section, we lay out the steps taken to identify loaded trips in the AIS data. These loaded trips form the basis of our shipping dataset.

3.4.1 Trips

In order to identify loaded trips, we must first recover the *trips* undertaken by each tanker, which are defined as the voyages between two consecutive departures from ports. Specifically, each trip consists of four stages: (1) departure from a port, (2) sailing at high speed, (3) arrival at a (different) port, and (4) staying in-port. To distinguish between these four stages, we must pin down a ship's moving status and differentiate between sailing and staying in-port. We draw on industrial practice and apply two criteria for labeling a ship as staying in-port: Its speed is less than one knot, and the geodesic distance between the ship's position and any port nearby is less than 50 kilometers. Subsequently, we obtain a time series of port calls (including port location and time stamp), and recover the trips in this manner.[12]

3.4.2 Loaded and Unloaded Trips

Next, we leverage the draught information provided by the AIS in order to distinguish between *loaded* and *unloaded trips*.[13] The draught measures the vertical distance

[12]In addition, we exclude trips that (1) are at the beginning/end of the sampling period, due to their potential incompleteness; (2) are discontinuous, in the sense that there exists a significant gap of no less than six days between two consecutive AIS observations; or, (3) have the same departure and arrival ports. The last criterion is set so as to account for occasions when a vessel visits the same port terminal twice during the same port call due solely to logistical reasons (e.g., maintenance need, cargo/tank capacity availability).

[13]It is noted that the draught information is manually recorded by ship crews, and thus may be subject to pervasive measurement errors compared to automatically recorded information, such as

from the bottom of a vessel's keel to the water's surface. It tells how deep the ship is sitting in the water, and reflects how much cargo the ship is carrying. A high and/or increasing draught indicates that a large and/or increasing proportion of the tanker's hull is becoming submerged, which implies that it is either loaded or in the process of loading oil.[14]

Taking advantage of such draught information, we identify loaded trips in two steps. The first step is to label the ship as loading when the trip draught increases (and unloading when the trip draught decreases) in comparison to its last trip draught. The peak in trip draught corresponds to the loaded state, whilst the trough corresponds to the unloaded state. We classify those readings in the middle as the partially loaded state. This single step does not completely solve our problem, however, due to gaps in the AIS data that inevitably render identification solely through variations in trip draught unreliable. Hence, for each ship's unlabeled observations that arise from data gaps, we apply the Gaussian mixture model to determine their respective states. Specifically, we first calculate the draught ratio of a trip as the ratio of the trip draught to the vessel's design draught.[15] Next, we draw the frequency distribution histograms of draught ratios for LR1 and LR2 ships, respectively, in Figure 3.1. In both cases, two distinctive peaks can be observed, one close to the ballast draught and the other close to the design draught; these correspond to the unloaded and loaded states, respectively. In between, there is a less obvious peak, which represents the partially loaded state that occurs less commonly during the sampling period. Finally, we adopt the Gaussian mixture model as the probability density function of the underlying data in order to take

speed and position. The errors can stem from typos, missing values, and updating time lags. To remedy this issue, we take the last observed draught before the ship enters its port of destination as the voyage draught of the trip. The reason is that, presumably, the crews will be more careful in reporting such information to port authorities before arrival in order to make sure that the port depth is adequate.

[14]As illustrated in Adland, Jia, et al. (2017), certain light product cargoes can be volume constrained rather than DWT constrained, such that vessel draught may not be indicative of a part-laden condition. This unavoidably leads us to falsely label some loaded trips as unloaded. This effect should be limited in principle, however, since the majority of trips identified in our dataset are labeled using the variations in trip draught rather than the absolute value. That is, as long as there is a loading/unloading process (as reflected in the variations in trip draught), loaded/unloaded trips can still be labeled accordingly.

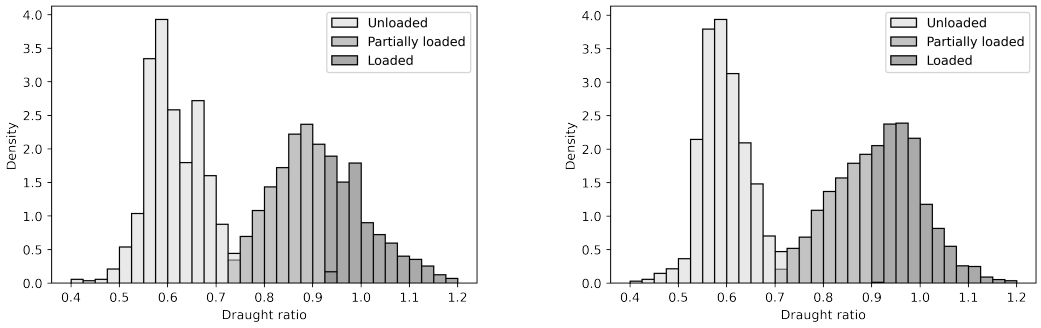
[15]Information on each vessel's design draught is obtained from the Lloyd's List ship database.

advantage of its ability to account for multi-peak distributions. In our setting, the model is applied to identify peaks in the distribution of trip draughts, and those peaks then represent the cut-offs that we apply when determining a ship's state. By denoting the ship's draught ratio during trip i as d_i and the cut-off draught ratios as d_U, d_{PL}, and d_L, respectively, a ship's state during trip i can be defined as follows:

$$\text{state} = \begin{cases} \text{unloaded}, & \text{if } d_i < \frac{d_U + d_{PL}}{2}; \\ \text{partially loaded}, & \text{if } \frac{d_U + d_{PL}}{2} \leq d_i < \frac{d_{PL} + d_L}{2}; \\ \text{loaded}, & \text{if } d_i \geq \frac{d_{PL} + d_L}{2}. \end{cases}$$

Figure 3.1: Draught ratio and state classification

(a) LR1 **(b)** LR2



Notes. The above figures plot the frequency distribution histograms of draught ratios for LR1 and LR2 tankers, respectively. The state classification of each level of draught ratio (i.e., unloaded, partially loaded, or loaded) is highlighted in different colors, based on the cut-offs derived from the Gaussian mixture model.

Consequently, we can pin down each ship's loading status and identify a loaded trip when it transitions through the partially loaded (if any) and loaded states over consecutive days. This gives us a total of 23,729 loaded trips (with non-empty and sensible information) taken by the 752 coated product tankers from 2017-2020.[16]

3.5 Estimation of Cargo Types

With the loaded trips identified, we can now recover the type of oil cargo carried by a ship in a loaded trip. This is critical, as it directly implicates the market

[16]In our analysis, we apply standard sample selection criteria by keeping (1) loaded trips with non-empty information on the input variables considered in the RF prediction model (see Section 3.5 for details), and (2) those with input variable values that fall within the 2.5th and 97.5th percentile range.

switching behavior of product tankers and, subsequently, the quantification of crude, dirty refined product, and CPP trades, respectively. To achieve this goal, we propose a novel two-step method, each step of which is illustrated in detail in the following sections.

3.5.1 Applying Existing Fixtures

In the first step, we match the empirically available fixtures to their corresponding, loaded trips in the AIS using the information on laycan dates. Such a strategy follows the existing literature on the matching of AIS and fixture data (Prochazka et al., 2019; Regli & Nomikos, 2019). Our fixture data include a total of 8,774 transactions for coated product tankers during the sampling period. We are not able to match all the records to the AIS data, however, because not every reported fixture is eventually realized. Hence, we consider only realized fixtures, by verifying whether a loaded trip identified in the AIS takes place within the stated laycan period, with a tolerance of $\pm$ 3 days. As a result, out of the 23,729 loaded trips identified in the AIS, we are able to determine a cargo type for 4,748 of them. Specifically, 1,568 (33.0%) are classified as crude oil trips; 692 (14.6%) are classified as dirty refined product oil trips; the remaining 2,488 (52.4%) are classified as CPP trips.

3.5.2 Applying the Random Forest Prediction Model

The second step involves the application of a supervised machine learning technique in order to predict the cargo types of the remaining, unclassified loaded trips. In particular, we opt for a random forest (RF) prediction model (Breiman, 2001). It has two stages; First, we use the RF model to distinguish between CPP and DPP trips made by the product tankers during the sampling period. Then, for those DPP trips, we further distinguish between crude and dirty refined product oil trips. Technically, the RF model is a powerful modification of bagging (also known as bootstrap aggregation), which builds a large collection of uncorrelated classification trees and then averages them.[17] With its ability to handle potential

[17]Details of the RF model, including the classification tree, are discussed in Section 3.5.2.2.

non-linear relationships between input variables, immunity to over-fitting, and high tolerance for outliers and noise within the data, the RF model is a natural choice for binary classification problems such as ours.

We consider several data sources in order to build the set of input variables for our RF prediction model. The primary source is still the AIS data, as described earlier. To obtain the vessel-specific information, we use the ship database from Lloyd's List Intelligence, which includes detailed information on the vessel's IMO number, year of construction, size, etc. Freight earnings are obtained from the Baltic Exchange. For the CPP shipping market, the Baltic Clean Tanker Index TC1-Time Charter Equivalent index (BCTI TC1-TCE) is used, as it tracks the dollars-per-day earnings for an average shipowner carrying 75,000 metric tons of CPP from Ras Tanura, Saudi Arabia to Yokohama, Japan. The route from Ras Tanura to Yokohama is considered as one of the major global shipping routes for CPP. Meanwhile, to calculate the average daily earnings for a shipowner in the DPP market, we apply the Baltic Dirty Tanker Index Aframax Time Charter Equivalent index (BDTI Aframax TCE), which is an average of the time charter equivalent earnings in six major, global Aframax shipping routes for DPP. Finally, we take advantage of the forward curves published by the Baltic Exchange for oil tankers in order to pin down the forward yield differences.

3.5.2.1 Input Variable Selection

The rational extraction of input variables provides the foundation of a machine learning model in terms of its predictive performance, as the combination of input variables essentially dictates the theoretical upper bound of predictive accuracy (Z. Liu et al., 2019). Therefore, based on extensive consultation with industry experts (who possess rich domain knowledge of this field), we have selected eleven input variables that, in principle, have significant power in differentiating between CPP and DPP trips in our first stage of cargo estimation.[18] These input variables

[18]The experts work as market analysts and brokers in a leading product tanker brokering company, and each has accumulated more than ten years of working experience. In our consultations with them, we asked three major questions related to the theme of our paper: (1) How significant is the role played by coated product tankers in the transportation of crude oil around the globe?

can be broadly grouped into three categories, namely, geography, market, and shipowner/vessel/operation heterogeneity; their summary statistics are presented in Table 3.1. In the following paragraphs we provide a detailed description of each input variable and our rationale behind its selection.

Table 3.1: Summary statistics of input variables (1st stage)

Input variable	**Unit**	**Mean**	**S.D.**	**Median**	**Min.**	**Max.**
Geography						
Trip distance	Nautical mile	3486.5	2964.8	2748.6	60.1	12410.6
Proportion of CPP trips from the origin country	Percent	50.5	33.4	50.5	0.0	100.0
Proportion of CPP trips to the destination country	Percent	53.7	30.2	54.3	0.0	100.0
Market						
Freight earnings spread	USD per day	931.4	14986.8	844.2	-37014.0	118119.5
Short-term forward yield differences	N/A	0.03	0.11	0.03	-0.41	0.33
Medium-term forward yield differences	N/A	0.03	0.14	0.06	-0.48	0.32
Long-term forward yield differences	N/A	-0.10	0.26	-0.03	-0.74	0.35
Shipowner/vessel/operation heterogeneity						
Proportion of CPP trips by the shipowner	Percent	55.6	34.6	62.9	0.0	100.0
Vessel's type	N/A	0.53	0.50	1.00	0.00	1.00
Vessel's age	Year	12.0	4.4	12.0	4.0	33.0
In-port duration	Day	6.5	5.4	5.0	1.0	29.0

Notes. The summary statistics are calculated using the information sources listed in Section 3.5.2. Each vessel's type is labeled 0 if it belongs to the LR1 category, and 1 if it belongs to the LR2 category.

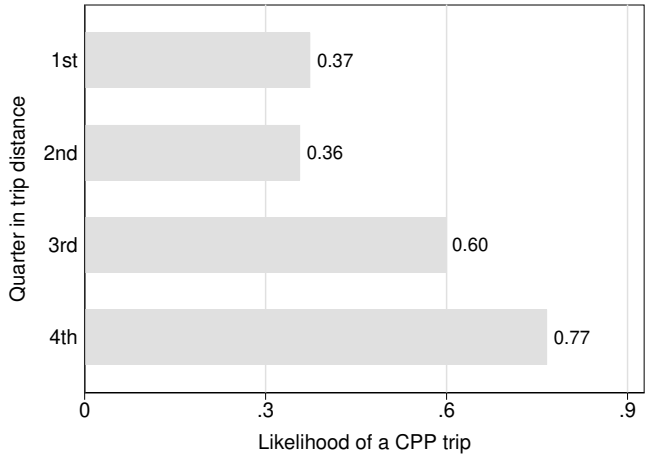
We first introduce three input variables related to the geography of the oil trade. To begin with, **trip distance** is defined as the sum of sailing distances between consecutive, reported positions in the AIS over the entire duration of a loaded trip. We include trip distance as an input variable because we observe a positive correlation between trip distance and the likelihood of a CPP trip in Figure 3.2. This pattern can be explained by the fact that the LR tankers considered in the dataset are considered to be not only the largest vessels used to carry CPP in long-haul voyages, but also the smallest vessels which can be used to carry crude oil; thus, they are most often used in intra-regional crude oil trade.[19] As a result,

(2) What are the factors that determine the decision by a product tanker to switch between CPP and DPP markets? (3) What are some of the major geopolitical events that underlie the drastic changes in the crude oil trade from 2017-2020? The last question is intended for an application of our prediction results to the study of the fluctuations in the global crude oil trade; this application is elaborated in Section 3.7.

[19]In practice, the refined product oil shipping market is considered to be more heterogeneous than the crude oil market, in that there are more types of refined petroleum products and a greater

the average trip distance of a CPP trip tends to be longer (though not uniformly) than that of a DPP trip in our shipping dataset.

Figure 3.2: The likelihood of a CPP trip and trip distance



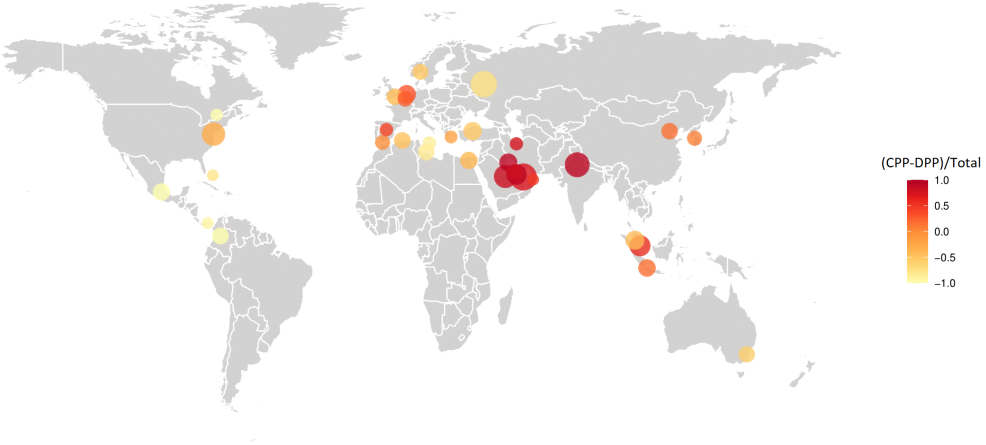
Notes. Each bar represents the proportion of CPP trips in a sample, with trip distance falling in each quarter of its distribution. For instance, the first bar indicates that, for loaded trips with a traveling distance that falls in the first quarter of its distribution, approximately 37% of them are carrying CPP; for those in the fourth quarter, roughly 77% are carrying CPP. The underlying sample of the figure includes those loaded trips with cargo types classified from the first step of the estimation (i.e., from applying existing fixtures).

In addition to trip distance, **the proportion of CPP trips in the country of origin** (of a loaded trip) is calculated as the proportion of CPP trips originating from a loaded trip's loading country, over the total number of loaded trips initiated from that country. Similarly, **the proportion of CPP trips in the country of destination** (of a loaded trip) is calculated as the proportion of CPP trips finishing in a loaded trip's discharge country, over the total number of loaded trips ending at that country. As shown in Figure 3.3, there are significant variations in the geographic concentration of the oil trade. On the one hand, most DPP trips originate from either the Persian Gulf or the Caribbean, and conclude in one of the major oil-importing countries in Europe (such as the U.K.) or Southeast Asia (Singapore, for example), because these oil-importing countries possess adequate oil refinery capacity. On the other hand, CPP trips often originate from the Persian

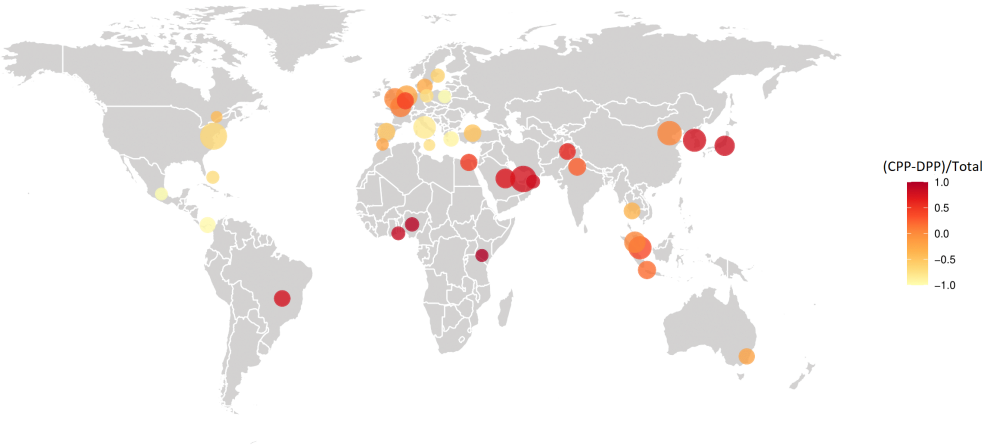
diversity of trading routes. As a result of the more homogeneous nature of crude oil shipping, the size of the vessels designed to carry crude oil (e.g., VLCC) is often much larger, and they tend to travel longer distances between continents than product oil tankers.

Figure 3.3: Geographic dispersion of oil trade

(a) Oil exports



(b) Oil imports



Notes. The first plot shows the geographic dispersion of CPP and DPP exports, while the second shows the dispersion of CPP and DPP imports during the sampling period from 2017W5-2020W48. To measure the concentration of the oil trade in a country, we use the difference between CPP and DPP trips over the total number of loaded trips originating from or ending at that country; a positive (negative) ratio indicates that a country is mostly exporting or importing CPP (DPP). The concentration is also differentiated by colors along a spectrum. The total oil exports or imports of a country are measured by the size of its pie. The plots are created using the information regarding the countries of origin and destination that is available for 90% of the loaded trips with a cargo type classified from the first step of the estimation (i.e., from applying existing fixtures). This helps us to avoid inferences based on samples with too few loaded trips.

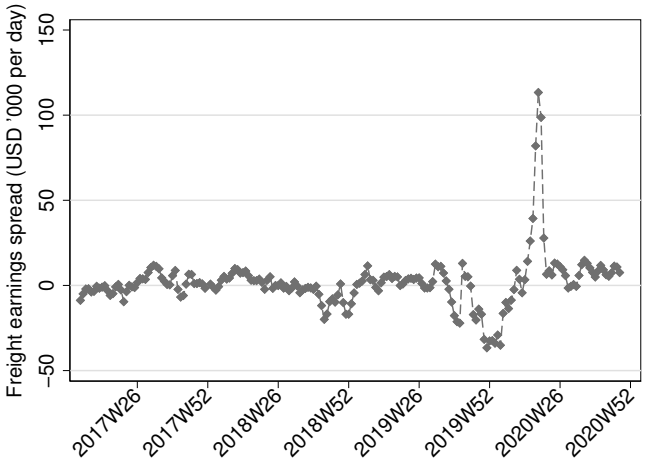
Gulf, Europe, or East Asia, but can end at almost any coastal, oil-importing country (except for those in South America and Africa). In addition, as is widely accepted in the maritime industry, whether a trip originates from east or west of the Suez Canal has significant predictive power in determining a vessel's cargo type: If the trip originates from east of the Suez Canal, there is a high probability that the ship is carrying CPP, and vice versa. These patterns can be clearly seen in Figure 3.3a. Hence, we include these two input variables in the estimation.

Next, we turn to the market-related input variables. **Freight earnings spread** is calculated by taking the first difference between the two earnings indices, BCTI TC1-TCE and BDTI Aframax TCE. Furthermore, by taking into account the fact that ships are often fixed in the AIS one to two weeks before their actual trips take place, we include the 14-day average spread prior to a ship's departure as one input variable. The reason we consider the freight earnings spread is straightforward, in that shipowners are profit-driven and tend to switch to the market with higher earnings. In Figure 3.4 we plot the weekly average freight earnings spread between the CPP and DPP markets observed during the sampling period. As demonstrated in the figure, the earnings spread remained relatively stable until the end of 2019, when it plunged to a historical low before soaring up to more than 100,000 USD per day in mid-2020. The spike was largely due to excess demand for CPP shipping services, as tankers were tied up as floating storage of crude oil following the contraction in demand caused by the COVID-19 pandemic (see Section 3.7 for a detailed discussion of this phenomenon). These short-lived (but unprecedentedly high) earnings provided strong incentives for tankers to search for delivery jobs in the CPP market; therefore, we observe a sharp increase in tanker utilization in the CPP market during the same period.[20]

Furthermore, **short-term, medium-term, and long-term forward yield differences** are calculated using the forward freight prices of CPP and DPP. These differences capture the extent to which the forward curve of CPP is trending upwards

[20]Tanker utilization is defined as the proportion of product tankers that are loaded with oil cargoes in a shipping market. See Bai and Li (2021) for a detailed discussion of the dynamics of tanker utilization from 2017-2020.

Figure 3.4: Freight earnings spread



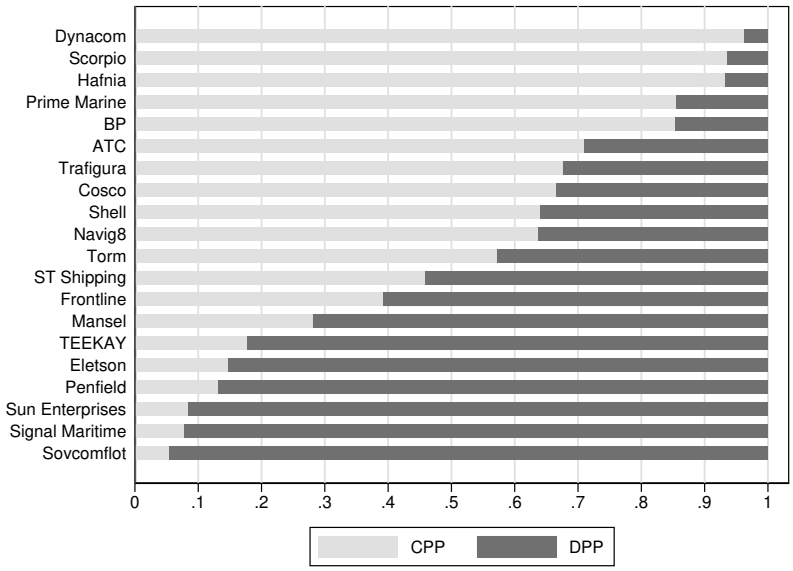
Notes. The freight earnings spread is calculated as the first difference between the average earnings index in the CPP market (BCTI TC1-TCE), and that in the DPP market (BDTI Aframax TCE) between 2017W5 and 2020W48.

compared to that of DPP, and essentially reflect the prospective earnings, market sentiments, and expectations in the oil markets. Thus, assuming that shipowners are forward-looking, we expect that they will be more likely to opt for CPP jobs when the forward yield differences are high. To calculate such differences, let x_0 and $\{x_t\}_{t\in[1,5]}$ denote the current freight price of CPP, and its forward freight prices within five months after the date recorded in the AIS two weeks before the actual trip. Likewise, define y_0 and $\{y_t\}_{t\in[1,5]}$ as the current and forward freight prices of DPP, respectively. The values of x are estimated using the mean forward freight prices of two major CPP shipping routes, that is, TC5 (55,000mt, Middle East to Japan) and TC15 (80,000mt, Naphtha Skikda to Chiba). The values of y are estimated using the mean forward freight prices of four major DPP shipping routes, namely, TD7 (80,000mt, North Sea to Continent), TD8 (80,000mt, Kuwait to Singapore), TD17 (100,000mt, Baltic to UK-Continent), and TD19 (80,000mt, Cross-Mediterranean). Having estimated the values of x and y, the short-, medium-, and long-term forward yield differences can be calculated as follows:

$$Diff_t = \begin{cases} \frac{(x_1-x_0)/x_0+(x_2-x_1)/x_1}{2} - \frac{(y_1-y_0)/y_0+(y_2-y_1)/y_1}{2}, & \text{for } t = ST; \\ \frac{(x_3-x_2)/x_2+(x_4-x_3)/x_3}{2} - \frac{(y_3-y_2)/y_2+(y_4-y_3)/y_3}{2}, & \text{for } t = MT; \\ \frac{x_5-x_4}{x_4} - \frac{y_5-y_4}{y_4}, & \text{for } t = LT. \end{cases}$$

We also introduce four input variables that capture the heterogeneity among shipowners, vessels, and shipping operations, respectively. **The proportion of CPP trips carried by the shipowner** (during a loaded trip) is calculated as the proportion of CPP trips that a loaded trip's designated shipowner engages in, over the total number of loaded trips completed by that shipowner. We select this input variable because we observe in the shipping data that certain shipowners mostly carry only one type of oil product during the sampling period. This is potentially due to a longstanding relationship between the shipowner and oil exporter, or to the shipowner's risk tolerance. For instance, in Figure 3.5, it is seen that Dynacom primarily transports CPP, while Sovcomflot mostly transports DPP.

Figure 3.5: Proportions of CPP (DPP) trips completed by the top 20 shipowners



Notes. The plot illustrates the proportion of CPP (DPP) trips completed by a shipowner among all its loaded trips, the cargo type of which has been classified from the first step of estimation (i.e., from applying existing fixtures data). The top 20 shipowners are selected based on the fact that they account for approximately 60% of the loaded trips identified in the shipping dataset. The shipowners in our context refer to those commercial vessel operators in the international shipping market.

In addition, we consider the technical specifications of a product tanker, as we acknowledge that certain types of vessel are more technically suitable for CPP transportation, or are preferred by charterers that trade CPP. Hence, we include the **vessel's type** as an input variable; this variable is labeled as 0 (1) when the

tanker belongs to the LR1 (LR2) category. Moreover, we gather information on each **vessel's age** by taking the first difference between the start year of a loaded trip and the construction year of the designated product tanker. This follows from the observation that tankers employed in the CPP market are, on average, younger than those in the DPP market. This is because many charterers that trade CPP set age limits on tankers, and prefer those with less than 15 years old due to the better maintenance of their onboard oil storage facilities.

Finally, we apply **in-port duration** in order to succinctly capture the heterogeneity of shipping operations. This duration is given by the total amount of time a ship spends in-port during a loaded trip. The rationale for its inclusion is that, as the loading and discharging operations are different for CPP and DPP, the time required to perform such operations should (in principle) vary.

After applying the aforementioned eleven input variables to the first stage of cargo type estimation (that is, distinguishing between CPP and DPP), we re-apply seven of them in the second stage of cargo type estimation in order to further differentiate between crude and dirty refined product oil trips among those estimated DPP trips.[21] Specifically, for each estimated DPP trip, we calculate its trip distance, the designated vessel's type and age, and the corresponding in-port duration. Moreover, we calculate the proportions of crude oil trips in the countries of origin and destination for an estimated DPP trip, along with the proportion of crude oil trips carried by the designated shipowner. We present the summary statistics of these calculations in Table B.1 for reference.

3.5.2.2 Estimation Details

After describing the input variables to be applied in the RF prediction model, we follow Breiman (2001) and implement our model as follows:

[21]The predictive performance of our RF model in the second stage of cargo type estimation is somewhat limited by the fact that we are not able to consider the market-related input variables, as we do not have data on the freight earnings or the forward freight prices of crude and dirty refined product oil. See Section 3.6.2 for further details.

1. Let B^{1st} denote the number of classification trees to be grown in the first stage of cargo type estimation. For tree $b = 1$ to B^{1st}:

   (a) Draw a bootstrap sample $\boldsymbol{Z}^{1st}$ of size N^{1st} from the training set;

   (b) Grow a classification tree T_b^{1st} with the bootstrap data, and recursively implement the following steps for each terminal node of the tree until some minimum node size n_{min}^{1st} is reached (see Figure 3.6 for an illustration):

      i. Select m^{1st} variables randomly from the eleven input variables described in Section 3.5.2.1;

      ii. Pick the best splitting variable and split point for the m^{1st} variables;

      iii. Split the terminal node into two successor nodes;

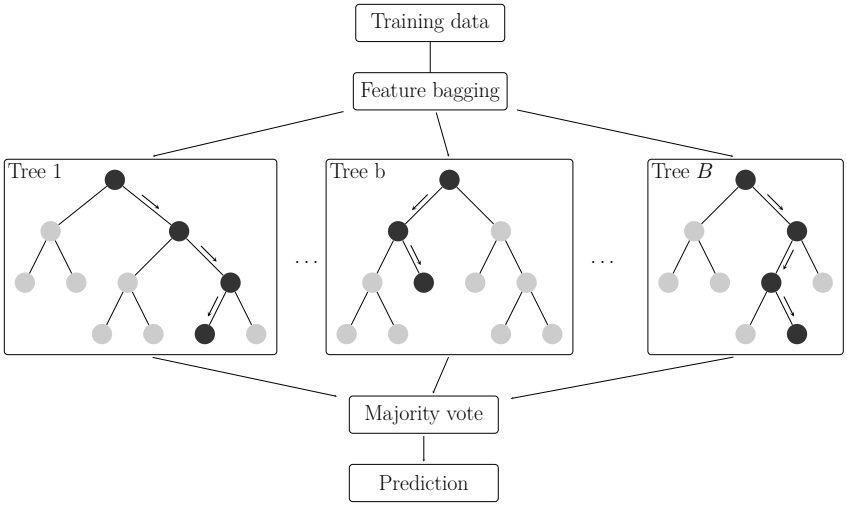
   (c) Let $\hat{C}_b^{1st}(x)$ denote the predicted class of sample x from the b-th classification tree, i.e., $\hat{C}_b^{1st}(x) \in \{CPP, DPP\}$. The predicted class of sample x is thus given by $\hat{C}^{1st}(x) = Majority(\hat{C}_1^{1st}(x), \ldots, \hat{C}_{B^{1st}}^{1st}(x))$;

2. Collect the set of estimated DPP trips from the first stage, and re-apply the RF prediction model in order to differentiate between crude and dirty refined product oil trips, i.e., $\hat{C}_b^{2nd}(y) \in \{crude, dirty\ refined\ product\}, \forall b \in B^{2nd}$. The predicted class of sample y is thus given by $\hat{C}^{2nd}(y) = Majority(\hat{C}_1^{2nd}(y), \ldots, \hat{C}_{B^{2nd}}^{2nd}(y))$.[22]

For our RF prediction model, the training set accounts for 80% of the loaded trips with a cargo type classified from the first step of the estimation (that is, from applying existing fixtures data), while the other 20% constitute the test

[22]The key underlying mechanism embedded in the formation of classification trees is recursive binary partition. We split the input variable space into a set of non-overlapping, multi-dimensional rectangles, and then fit a fixed class (say, the majority class) to each one. For instance, we consider a classification problem with discrete response y and input variables $(x_1, \ldots, x_J)$. When growing a classification tree, the partition is performed recursively. First, we select one of the input variables x_j and the corresponding split point s to split the J-dimensional input variable space into two subspaces. One subspace contains all the observations in which $x_j \leq s$, and the other captures those in which $x_j > s$ (e.g., whether or not the vessel's age is less than 15 years). Subsequently, we split each subspace in a similar manner, by choosing the splitting input variable and the corresponding split point. This operation is repeated until some minimum node size is reached (i.e., the minimum number of observations in a single multidimensional rectangle). Figure 3.6 illustrates this process by plotting several binary trees: In each tree, the full input variable space is the root node, and each split divides the internal node into two successor nodes.

Figure 3.6: Basic structure of an RF prediction Model



set. Throughout the process, we balance the class distribution within the splits when sampling the dataset. Furthermore, we perform parameter tuning and model training on the training set, and the model's predictive performance is evaluated using its predictions about the test set.

There are six, tunable hyper-parameters in our RF prediction model, namely, m, N, n_{min}, B, *replacement*, and *splitting rule*. Their values in both stages of the cargo type estimation process are summarized in Table 3.2. To begin with, lower values of m and N tend to result in fewer correlated classification trees, thus contributing to reduced variance during aggregation. Nevertheless, the performance of a single tree is likely to be impaired in this way, since it is built on a set of sub-optimal variables (Probst et al., 2019). As a result, m and N should be carefully tuned in order to handle such a trade-off. The minimum node size n_{min} determines the depth of a classification tree, and Lin and Jeon (2006) point out that the RF model's performance can be improved when n_{min} is tuned with care. Therefore, in our setup, we tune m, N, and n_{min} through cross-validation, so as to increase the overall predictive performance of our RF prediction model.[23]

[23]One problem that we encounter in the second stage of cargo type estimation is related to the imbalanced sample. Specifically, in the training set, the number of crude oil trips is approximately twice that of dirty refined product oil trips. This inevitably leads to a poor performance of our RF model when predicting the cargo type of a loaded trip that actually carries dirty refined product oil. To address this problem, we augment the cross-validation process by applying the synthetic minority over-sampling technique (SMOTE) when tuning the hyper-parameters of the

By contrast, the three remaining hyper-parameters are not tuned in our setup, on account of the properties of the RF model and the relevant characteristics of the selected input variables. Specifically, Probst and Boulesteix (2017) argue in favor of setting B as large as computationally possible, and recommend not tuning it at all. In terms of *replacement* and the *splitting rule*, we believe it is justifiable not to tune either of these terms, as the input variables in our context are all numeric by nature. Here, we depart from Strobl et al. (2007), who claim that the variable selection procedure may be biased when data of various kinds are dealt with in the process of bootstrap sampling with replacement, and from Probst et al. (2019), who argue that the splitting rule of minimizing Gini impurity has a proclivity to select variables with more possible splits.

Table 3.2: Hyper-parameters in the RF model

Hyper-parameter	**1st stage**	**2nd stage**
m	5	3
N	0.7n	0.8n
n_{min}	1	3
B	1000	1000
replacement	true	true
splitting rule	minimizing Gini impurity	minimizing Gini impurity

Notes. The table lists the values of six, tunable hyper-parameters in our RF prediction model. Details of these hyper-parameters are illustrated in the main text. SMOTE is applied to tune m, N, and n_{min} in the second stage of cargo type estimation, and the number of nearest neighbors and the duplicated size of the minority class are set at 3 and 1, respectively. n is the sample size of the training set.

3.6 Evaluation of Prediction Results

3.6.1 CPP vs. DPP Trips

In this section, we apply several metrics to assess the predictive performance of our RF model in the first stage of cargo type estimation, in which we seek to differentiate between CPP and DPP trips. We first examine the model's predictive

RF model in the second stage of cargo type estimation. SMOTE is an over-sampling technique that is frequently applied to tackle the imbalanced sample problem in supervised machine learning (Chawla et al., 2011). When predicting the cargo types for the test set, we also adopt SMOTE to pre-process the training set, and to train the RF model accordingly.

accuracy (ACC) score. The ACC score is calculated as the proportion of cargo types that the model correctly classifies, and is defined by:

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \times 100\%,$$

where TP (true positive) and TN (true negative) are the numbers of cargo types correctly classified, and FP (false positive) and FN (false negative) are the numbers of cargo types incorrectly classified. In the first stage, by referring to the confusion matrix presented in Table 3.3, we see that the out-of-sample ACC for the test set stands at 93.7%. In addition, the value of sensitivity, $TP/(TP + FN)$, is 94.2%, indicating that out of all CPP trips, 94.2% of them are correctly classified. The value of specificity, $TN/(TN + FP)$, is 93.1%, implying that 93.1% of all DPP trips are correctly classified. Such high values of ACC, sensitivity, and specificity highlight the power of our RF model in differentiating between CPP and DPP trips.

Table 3.3: Confusion matrix for the test set (1st stage)

		Actual cargo type	
		CPP	DPP
Predicted cargo type	CPP	468 (TP)	31 (FP)
	DPP	29 (FN)	420 (TN)

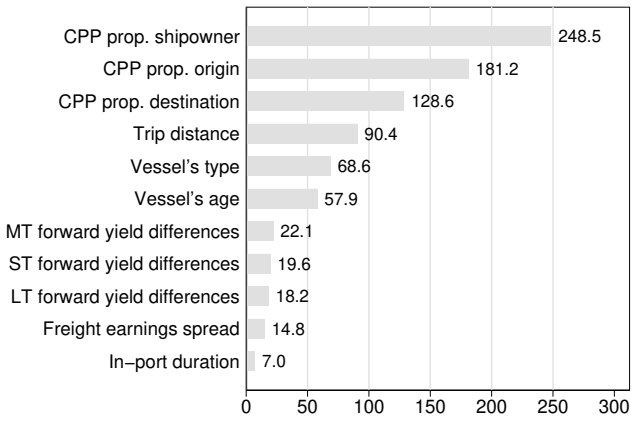
Notes. In the confusion matrix, the rows and columns represent the predicted and actual labels, respectively. In addition to the incidences of each category, we write the corresponding classifications in parentheses.

The predictive performance of the model is further evaluated by the receiver operating characteristic (ROC), or the ROC curve. The ROC curve is created by plotting the true positive rate against the false positive rate at various threshold settings. The area under the ROC curve (AUC) measures the capacity of a prediction model to distinguish between classes for a particular classification problem. The AUC ranges from 0.5 to 1, and a larger AUC indicates a better predictive performance in general. The AUC in the first stage is 0.98, which implies that there is a 98% chance that the model correctly distinguishes between CPP and DPP trips.

The relative importance of each selected input variable to the classification of cargo types can be further determined by the mean decrease accuracy (MDA)

value, which indicates a variable's overall discriminating power in the underlying classification problem. Variables with a higher MDA are thus of greater significance in classifying the data into correct classes. Figure 3.7 plots each input variable's MDA. As seen in the figure, the most critical variables are the proportion of CPP trips carried by the shipowner, the proportion of CPP trips from the country of origin, and the proportion of CPP trips undertaken to the country of destination of a loaded trip. These are followed by trip distance, the vessel's type and age, and the medium-, short-, and long-term forward yield differences. The freight earnings spread and in-port duration display less predictive power than the other variables.

Figure 3.7: MDA values of the input variables (1st stage)



Notes. Based on their MDA values, the selected input variables are ranked from greatest to least relative importance in classifying loaded trips into the correct classes.

A closer examination of the MDA values indicates that the fluctuations of shipping markets do not necessarily impact the decisions of shipowners, as the forward yield differences and freight earnings spread are much less important to the classification of cargo types than many of the other input variables. In particular, the greater importance of a loaded trip's designated shipowner and its countries of origin and destination implies that the shipping patterns of global oil trade are rather sticky. This is consistent with both the literature and empirical evidence. For natural resources like crude oil, their origins and destinations are remarkably fixed, leading to sticky trade flows in this type of market. This is largely due to the fact that countries with abundant resources naturally tend to export, while

those without sufficient resources must import inputs for which no close substitutes are available.[24] For other types of product oil, although their trade flows are less sticky than those of crude oil, their trade patterns are also remarkably fixed and predictable (Bai & Lam, 2019).

In addition to the above measures, we also evaluate the overall performance of our prediction model against a baseline model that only considers the technical specifications of a product tanker (i.e., the vessel's type and age). The out-of-sample ACC, sensitivity, specificity, and AUC of such a baseline model in the first stage of cargo type estimation are 76.6%, 78.3%, 74.7%, and 0.81, respectively, all of which values are significantly lower than those associated with the full-fledged model. Such comparisons thus clearly indicate the importance of geography-, market-, and shipowner-related factors in determining the global distribution of CPP and DPP trips.

Finally, to check the robustness of our prediction results, we introduce three alternative prediction models (namely, logistic regression, support vector machine (SVM), and neural network), and compare their predictive performance to that of RF across the evaluation metrics of ACC, Sensitivity, Specificity, and AUC.[25] Table 3.4 presents a summary of the prediction results on the test set using each of the aforementioned classifiers. The results demonstrate that the RF model achieves the best overall predictive performance, even though the differences are only marginal.[26] This is reasonable, as the bootstrap aggregating strategy makes RF a powerful supervised learning model when some input variables do not explain the response variable very well. In terms of the performance of logistic regression, although it helps us verify the relationship between the input and response variables (see Table B.3), its simplicity and under-fitting issues make it perform relatively poorly in terms of predictive accuracy. As for SVM, the difficulty of choosing the best

[24]Our finding is also consistent with Prochazka et al. (2019), as they assert that the pattern of international oil trading is steady and well-defined, with a clear fronthaul-backhaul structure.

[25]In Appendix B.2, we briefly discuss the underlying mechanism embedded in each of the three alternative prediction models.

[26]Such a robust result also indicates that we may have achieved the theoretical upper bound of predictive accuracy, which is largely determined by the rational extraction of the input variables.

kernel function limits its performance. Lastly, since there are only eleven input variables, the advantages of SVM and neural network in handling high-dimensional data cannot be fully exploited in the estimation of cargo types.

Table 3.4: Summary of the prediction results on the test set (1st stage)

Classifier	**ACC**	**Sensitivity**	**Specificity**	**AUC**
Random forest	93.7%	94.2%	93.1%	0.980
Logistic regression	92.7%	92.0%	93.6%	0.974
SVM	93.6%	93.4%	93.8%	0.973
Neural network	92.8%	94.0%	91.6%	0.972

Notes. ACC measures a model's overall predictive accuracy, and is calculated as the proportion of cargo types that the model correctly classifies. Sensitivity measures the proportion of CPP trips that are correctly classified, while specificity measures that of DPP trips. AUC measures the ability of a model to distinguish between CPP and DPP trips.

3.6.2 Crude vs. Dirty Refined Product Oil Trips

In this section, we apply the same metrics to assess the predictive performance of the RF model in the second stage of cargo type estimation, in which we further differentiate between crude and dirty refined product oil trips among those DPP trips estimated in the first stage. By referring to the confusion matrix in Table 3.5, we summarize that the out-of-sample ACC, sensitivity, specificity, and AUC of our RF model stand at 76.5%, 60.9%, 83.4%, and 0.77, respectively. In addition, as seen in Table B.2, the three alternative classifiers produce almost the exact same prediction results. We attribute the relatively weak predictive performance of our RF model (as compared to its performance in the first stage) to the problem of an imbalanced sample, as the number of crude oil trips is approximately twice the number of dirty refined product oil trips in the training set. Finally, by comparing the MDA values of the selected input variables in Figure B.1, we find that (consonant with our observation in the first stage) geography still plays a more significant role in determining the distribution of crude and dirty refined product oil trips than do the other technical factors.

Table 3.5: Confusion matrix for the test set (2nd stage)

		Actual cargo type	
		Dirty refined product	Crude
Predicted cargo type	Dirty refined product	84 (TP)	52 (FP)
	Crude	54 (FN)	261 (TN)

Notes. In the confusion matrix, the rows and columns represent the predicted and actual labels, respectively. In addition to the incidences of each category, we write the corresponding classifications in parentheses.

After estimating the cargo types using the RF prediction model, we determine that, out of the 23,729 loaded trips identified in the AIS from 2017-2020, 5,923 (25.0%) can be classified as crude oil trips, 5,130 (21.6%) can be classified as dirty refined product oil trips, and the remaining 12,676 (53.4%) can be classified as CPP trips. More importantly, the 5,923 crude oil trips represent approximately 8% of the global seaborne crude oil volume traded between 2017 and 2020, indicating the vital role of coated product tankers in the international transportation of crude oil.

3.7 Application of Prediction Results

In this section, we apply the shipping dataset developed in the previous sections, and perform a quantitative exercise that studies fluctuations in the global seaborne crude oil trade volume carried by the LR tankers from 2017-2020. This historical analysis serves as an external validation of our estimation strategy, and showcases the model's ability to identify major oil trade disruptions.

We first consolidate all the crude oil trips in the constructed shipping dataset, and calculate the ton-miles of each trip by multiplying the trip distance by the DWT of the designated product tanker. In maritime transportation, a ton-mile is defined as one ton of cargo transported one mile. It is a standard measurement of the overall demand for cargo shipping services, and fundamentally reflects the overall level of industrial activity (Dennis, 2004). In this paper, we approximate the trade volume of crude oil carried by LR tankers worldwide on a single day t by using the total, daily ton-miles of the crude oil trips that have already set out but

have not yet arrived on day t. This is reflected in the following formula:

$$trade_t = \sum_{i \in \{i | s_i \leq t, e_i \geq t\}} \frac{D_i \times DWT_i}{e_i - s_i + 1},$$

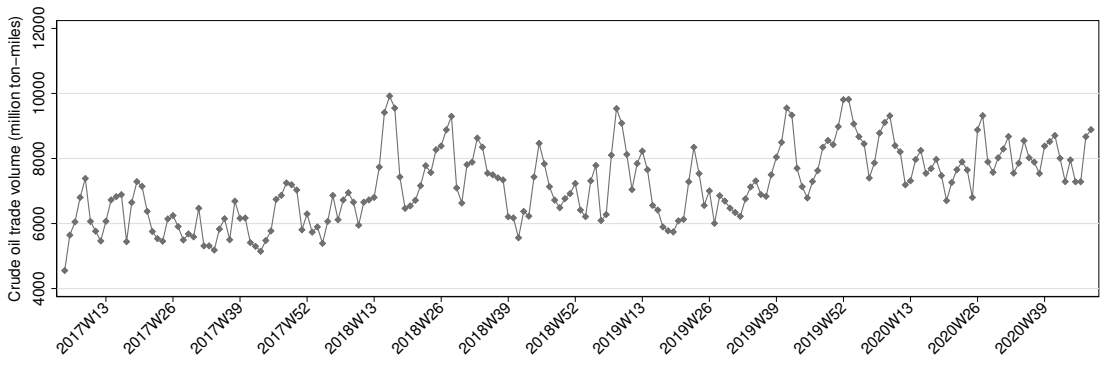
where $trade_t$ is the global crude oil trade volume carried by LR tankers on day t, and D_i, DWT_i, e_i, and s_i are the trip distance, DWT, the arrival date, and the start date of crude oil trip i, respectively.[27]

In order to illustrate these calculations, we add the total daily ton-miles in each week of the period 2017W5-2020W48 together, and plot in Figure 3.8 the raw weekly series of crude oil trade volume carried by LR tankers across the world. At first glance, the series fluctuates significantly during the sampling period, with several easily discernible peaks and troughs occurring amid recurrent seasonal variations. In order to remove these anticipated seasonal variations from the time series and study the impacts of unanticipated trade disruptions on the global crude oil trade, we follow the literature (Cleveland et al., 1990) and employ the seasonal-trend decomposition procedure based on loess (STL). In its application, STL decomposes a time series into three components (seasonal, trend, and remainder), and it is implemented in two looping operations. During each iteration of the inner loop, the seasonal and trend components are updated once, based on loess and on the robustness neighborhood weights provided by the former iteration of the outer loop. According to Cleveland et al. (1990), all the parameters in STL, except for the seasonal smoothing parameter n_s, can be determined straightforwardly. Nevertheless, as shown in Figure B.2, the choice of n_s exerts little influence on the results of our decomposition, as the remainder components extracted by using different values of n_s do not depart from each other significantly.

By applying a smoothing parameter $n_s = periodic$, we can plot the remainder component of STL, as seen in Figure 3.9. After removing the seasonal and trend components from the time series, this remainder series essentially shows us the unanticipated variations in the global crude oil trade volume carried by LR tankers.

[27]For simplicity's sake, we posit that a product tanker sails at the same constant speed during each of its loaded trips. Hence, the daily ton-miles of a trip can be easily approximated by dividing the trip's ton-miles by the number of days that the tanker travels.

Figure 3.8: Global crude oil trade volume carried by LR tankers from 2017W5-2020W48

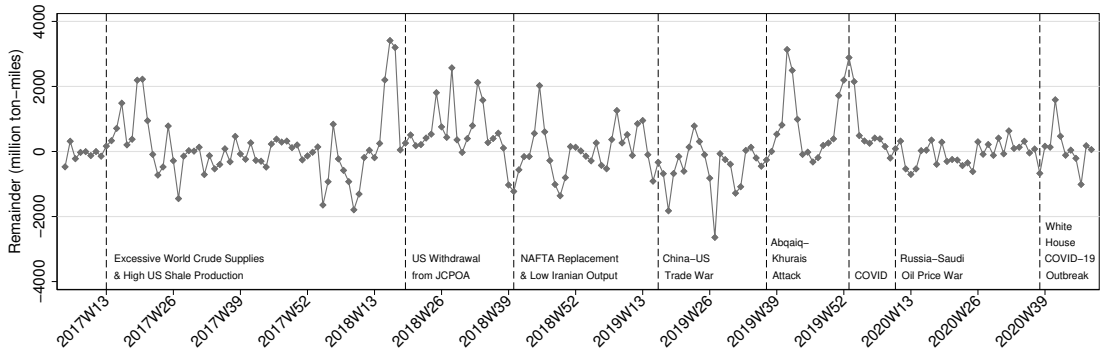


Notes. The global crude oil trade volume transported by LR tankers on a single day is approximated by the total daily ton-miles of the crude oil trips that have already set out but have yet to arrive on a given day. We add the total, daily ton-miles in a week together to determine the weekly ton-miles, and then we use this weekly series to approximate the trade volume of crude oil carried by LR tankers.

These fluctuations usually stem from exogenous trade disruptions caused by external events. By comparing our plot with the chronology of world oil markets, it turns out that our remainder series accords quite well with some key historical episodes. In early 2017, although the Organization of the Petroleum Exporting Countries (OPEC) agreed to curb crude supplies in a bid to end the worst oil crash in a generation, its attempts to bolster the market were thwarted by a quick rebound in U.S. shale production, and by signs that OPEC exports did not fall as quickly as expected, despite a high degree of member compliance with the mandated production cuts (Raval & Sheppard, 2017). This excess of oil supply translated to a 14% tumble in the West Texas Intermediate (WTI) crude oil futures price in the first half of 2017, and a corresponding increase in the global crude oil trade, the latter of which can be clearly seen in Figure 3.9.

Subsequently, in 2018 the oil market was perturbed by a series of geopolitical events. Notably, since the U.S. announced its withdrawal from the the Joint Comprehensive Plan of Action (JCPOA, also known as the Iran nuclear deal) and reinstated the economic sanctions targeting Iranian oil, Iran's crude oil exports declined significantly, from the peak of 2.5 million barrels per day in early 2018 to 0.6 million barrels per day at the end of 2018 (Landler, 2018). As Iran is one

Figure 3.9: Remainder component of STL with $n_s = periodic$



Notes. The figure plots the remainder component of an STL decomposition of the estimated global crude oil trade volume carried by LR tankers, with the seasonal smoothing parameter set at $n_s = periodic$. The vertical dashed lines indicate several major oil trade events that occurred during the sampling period, namely, the excess world supply of oil and high U.S. shale production in 2017W13; the U.S. withdrawal from the Joint Comprehensive Plan of Action (JCPOA, also known as the Iran nuclear deal) in 2018W19; the agreement to replace the North American Free Trade Agreement (NAFTA) and low Iranian oil production in 2018W40; the China-U.S. trade war in 2019W16; the Abqaiq–Khurais attack in 2019W37; the COVID-19 pandemic in 2020W1; the Russia-Saudi oil price war in 2020W10; and, the White House COVID-19 Outbreak in 2020W38.

of the major exporters of crude oil, its decreased exports delivered a significant shock to the oil transportation industry, as the crude oil trade was severely affected (U.S. Energy Information Administration, 2021).

Starting in 2019, the intense fluctuations of the crude oil trade are consistent with heightened uncertainty in global oil markets due to several interwoven factors. These included, but were not limited to, trade frictions, a weak international economy, and terrorism events. Specifically, the trough recorded between April and June 2019 is most likely to be associated with low demand in the international oil market owing to international trade disputes like the China-U.S. trade war (Kimball, 2019; Xia et al., 2019). On the other hand, the rise in the remainder series between September and October 2019 is probably due to a series of terrorism events.[28] For instance, the Abqaiq–Khurais attack occurred in September 2019, during which crude oil production facilities in Saudi Arabia were severely damaged by a drone attack, resulting in the near halving of its oil production. At the time, Saudi Arabia's

[28]This is because geopolitical news (including news of terrorist actions) can lead to a significant increase in the trading volume of crude oil, as the news can increase future uncertainty in the oil market (Brandt & Gao, 2019).

output represented about 5% of global oil production. The shock itself, together with its negative impacts on market sentiment due to increased concern about the oil supply, resulted in a sharp increase in the demand for oil and (correspondingly) for transportation services, the latter of which can be clearly seen from the rise in the crude oil trade during 2019W37-2019W44.[29]

Finally, we turn to examine the COVID-19 pandemic. Beginning from the end of 2019, despite massive demand destruction due to the pandemic, the demand for shipping services in the crude oil market actually rose. This was due to high levels of crude oil production in both Russia and Saudi Arabia that resulted in tankers being used as floating storage for the excess oil inventory (Camp et al., 2020). Following this peak, however, the demand dropped significantly, a continuous decline that can be seen in the remainder series covering the first half of 2020 in Figure 3.9. This can be explained as a result of skyrocketing freight earnings in the CPP market, which led tankers to switch markets (see Figure 3.4). Due to massive demand destruction after the lockdowns, oil refineries reduced their rates of operation, but not as swiftly as the drop in demand. Consequently, there was a large surplus of CPP. Domestic inventories quickly filled, and exporters used ships as floating storage, leading to excess demand for shipping services and high freight earnings in the CPP market.

The above results indicate that our method based on AIS data mining and machine learning has great potential for supervising and analyzing oil trade flows in real time. The resulting shipping data and the associated decomposition strategy enable relevant stakeholders to better engage in more targeted decision-making. For instance, shipowners could leverage the information provided by the model to identify profitable opportunities in a particular shipping market and make a more informed switching decision. Simultaneously, policymakers could apply it to quickly identify emerging trade flow risks and to craft sound economic policies in response.

[29]This is mostly related to the special role of Saudi Arabia in OPEC. It is by far the largest oil producer in OPEC, with enough capacity to function as a swing producer, if necessary, to balance the global market. For this reason, it is often thought of as "OPEC's de facto leader", a term which is routinely used in the financial press, e.g., https://www.ft.com/content/1f84e444-9ceb-11e5-8ce1-f6219b685d74 (Accessed 13 July 2021).

3.8 Conclusion

In this paper, we propose a big data approach to the classification of cargo types for coated product tankers, so as to estimate global crude oil trade volume in a more accurate and tractable manner. This approach overcomes existing difficulties faced by researchers in the estimation of oil trade, as it is built on an RF-based ensemble learning prediction model that utilizes the AIS data of coated product tankers. In this setup, the theory is well-established and straightforward for effective generalization, and the data are more accessible to the research community than is external information, such as line-up reports or shipping contracts.

Despite the encouraging performance achieved by the proposed approach, it is constrained by one technical difficulty, which is related to the fact that only product tankers with a well-maintained coating are able to achieve the cleanliness standard required by high-grade, refined product cargoes and successfully switch between the crude and refined product oil trades. This drawback, coupled with other commercial and organizational barriers illustrated in Adland, Hansson, et al. (2017), means that the selected input variables for our RF prediction model may not fully capture the actual occurrence of market switching behavior.

Finally, additional input variables (e.g., the segregation of upstream and downstream transportation activities or the shipowner-charterer relationship) could be incorporated to improve the performance of our prediction model once such data become available. Moreover, future research might consider incorporating more granular information, such as the terminal information of a vessel (when the exact terminal locations of a port can be identified from the AIS using a tailed algorithm, since manual labeling can be quite labor-intensive given the large number of ports involved). Our estimated oil trade flows can also be combined with those carried by crude oil tankers (e.g., VLCC) so as to provide a more comprehensive view of global crude oil trade flows.

4

The Congestion Effect of Oil Transportation and Its Trade Implications

4.1 Introduction

> *"Maritime transport is the backbone of global trade and the global economy, . . . , yet the vast majority of people are unaware of the key role played by the shipping industry, which is largely hidden from view."*
>
> — Ban Ki-moon, World Maritime Day 2016

Seaborne transportation is integral to global trade; it constitutes over 80% of total world trade by volume and more than 70% by value (United Nations Conference on Trade and Development, 2019). In particular, huge quantities of crude and product oil are transported by sea between production sites, refineries, and points of consumption due to limited pipeline infrastructure between trading partners.[1] Despite the indispensability of maritime transportation to worldwide trade, there is limited knowledge of how the shipping markets operate and how the trade shocks hitting these markets generate large aggregate fluctuations. In particular, the role of maritime transportation for the propagation of oil trade shocks in the world economy is still largely unknown.

[1]For reference, product oil refers to clean petroleum products (CPP), such as naphtha, gasoline, and jet fuel; they account for more than one quarter of total oil and petroleum product volumes traded worldwide (Trafigura, 2019).

Our paper is the first to study this dynamic, and our contributions include both a unique, granular shipping dataset of oil tankers, and a novel search and matching model of oil transportation that fits the data. We focus on the dynamic interactions between oil tankers and exporters following oil trade shocks that determine congestion in shipping markets. As we demonstrate, congestion is critical to the macro aggregates as it entails a trade-off between the volatility of the oil trade and final good production, and that of oil prices.

The application of quantitative research methods to the macroeconomic role of oil transportation has long been hindered, primarily due to the lack of appropriate shipping data. In order to study the micro behaviors of oil tankers, we need a high frequency, longitudinal shipping dataset that tracks each tanker's shipping status. Traditionally, such a quest has been challenging, since only aggregate trade or economic data have been available, and these are often of low frequency, delayed in publication, and inaccessible to the public. Given these limitations, we instead take advantage of modern Automatic Identification System (AIS) data that provide unprecedented information about the real-time movements of oil tankers, and subsequently construct a shipping dataset using state-of-the-art machine learning techniques. Leveraging this unique dataset, we establish three novel facts about the oil transportation industry. The first is related to the presence of shipping market congestion, which we show by estimating an aggregate matching function and finding that, for a one percent increase in the number of searching ships, the average contracting probability for ships decreases significantly by 0.6 percent, holding all else constant. The second finding is associated with the prevalence of market switching behavior by tankers. By calculating their market switching probabilities, we find that, on a weekly average basis, more than one third (one fifth) of unloaded ships in the crude (product) oil market that find oil cargoes to deliver are hired by product (crude) oil exporters. The third fact is linked to the propagation of oil trade shocks in the global economy through the transportation network. We show this process by linking the observed variations in tanker utilization in each oil shipping market (i.e., the proportion of tankers that are loaded with oil cargoes

in each market over the entire fleet of product tankers) to several major oil trade disruptions from 2017-2020, and find that they accord well with each other. These three facts are important, as they reflect that the shipping markets are congested, connected, and responding to oil trade shocks in a network structure.

Next, we develop a Diamond-Mortensen-Pissarides (DMP) random search model that accounts for these three novel facts and incorporates the interactions between oil exporters, tankers, and a final good producer.[2,3] There are two main innovations in the model. At its core, the fundamental contribution is the description of shipping market congestion. We set up the model in such a way that congestion arises when the shipping market becomes less tight, i.e., when the supply of oil transportation services exceeds their demand. In addition, the optimal searching behavior of oil tankers within and between markets depends crucially on how tight each shipping market is – the lower the tightness, the more congested the shipping market becomes, and hence the less likely the tankers will search for oil cargoes to deliver in such a market. The other innovation of our model is the linkage between the dynamics within different shipping markets and those associated with macro aggregates. Specifically, we establish that the oil trade depends on exogenous oil trade shocks (as reflected in the oil cargo delivered by each ship), together with the corresponding endogenous adjustment of the oil transportation network (as reflected in the number of ships transporting oil cargoes). As for oil prices, they are driven by the behaviors of ships between shipping markets, whereas the behaviors themselves

[2]We justify our choice of a DMP random search model as follows: First, a search and matching framework is ideal in our context because it can explicitly account for the state transitions of a ship that we observe in reality (i.e., unloaded or loaded). In addition, our finding of the presence of search frictions in the oil transportation industry (see Appendix C.1) also supports the application of such a framework. Next, we opt for a random rather than a directed search model, mainly due to its ability to explicitly model the negotiation process between ships and exporters. As we demonstrate, the negotiation process endogenizes the surplus sharing between a ship and an exporter in each shipping market, and the resulting differences in negotiated commissions rationalize the market switching behavior by ships. Moreover, the process of commission negotiation is consistent with industrial practice, as it is how spot freight rates are usually determined in reality.

[3]In what follows, we describe the model structure tailored to the oil transportation industry. However, as readers will see in the main text, the baseline model is developed in a setting with multiple commodity shipping markets, implying that our framework could be easily applied to other industries with similar features as what we see for oil transportation.

are influenced by shipping market congestion, among other primitives. Lastly, final good production combines the oil inputs, and thus inherits the variations therein.

We take advantage of available shipping data to tailor our empirical implementation of the model to match the micro behaviors of oil tankers. This involves first calculating various moments from the shipping data, and subsequently applying the Simulated Method of Moments (SMM) to match model moments to their empirical counterparts and determine several free parameters that cannot be calibrated externally. The specific SMM procedures we apply follow Chernozhukov and Hong (2003) and Lise (2013), in which a Markov Chain Monte Carlo (MCMC) estimation approach is undertaken. Using this approach, the model matches the selected data moments, including tanker utilization, shipping market shares, and the contracting and market switching probabilities of ships.

We use our empirically congruent model to investigate three important and related questions: (1) How does shipping market congestion within the oil transportation network adjust to exogenous oil trade shocks? (2) How does congestion affect the searching behaviors of ships within and between markets? (3) What are the implications of congestion for the oil trade, oil prices, and final good production? We study the first two questions by simulating the model and show that a positive oil trade shock favors a ship's search for a contract within the affected market, hence leading to a rise in tanker utilization. As more ships switch to the impacted market, however, the market becomes less tight until the supply of shipping services exceeds demand. As a result, the benefits of the shock (greater contracting probability and commissions) start to be depleted and shipping market congestion intensifies. The congestion then drives down the contracting probability for ships and paid commissions, leading to reductions in tanker utilization. In this way, excessive flows of ships into a market that receives a favorable trade shock are prevented, such that the fluctuations in tanker utilization are attenuated.

Subsequently, we illustrate the significant implications of shipping market congestion for the trade and price of oil and for the production of a final good through counterfactual exercises. First, by comparing our baseline setup to one in

which all the tankers are utilized to carry oil cargoes (hence congestion no longer manifests), we show that allowing for congestion is crucial, as it stabilizes more than one third of the volatility in the oil trade and in the final good production that would have occurred otherwise. The stabilization also comes with a more-than-tenfold increase in the volatility of oil prices, as they are directly determined by the dynamics of the shipping markets. Second, we consider another counterfactual exercise in which ships are not allowed to switch between markets, and show that the market switching behavior of ships enables asymmetric changes in congestion between shipping markets, with distinct implications for oil trade stabilization in each market. Lastly, by contrasting our baseline setup with one in which oil trade shocks no longer occur in shipping markets, we illustrate that it is the exogenous oil trade shocks that create most of the observed aggregate fluctuations, while the corresponding endogenous adjustment of the oil transportation network and the resulting congestion effect only play a minimal role in the process.

Related literature. Our study contributes to three strands of literature: (1) trade and transportation networks; (2) search and matching; and, (3) oil prices and shocks.

First, our paper links to the literature that studies the determination of trade flows in international trade (Allen (2014), Brancaccio et al. (2020), Eaton and Kortum (2002), and Wong (2018)). In particular, the pioneering work by Brancaccio et al. (2020) offers novel insights into the role of the transportation sector in attenuating differences in comparative advantages across countries, creating network effects on trade costs, and dampening the impacts of shocks on trade flows. Our setting also highlights the role of the oil transportation network in dampening the impacts of trade shocks on the oil trade; however, the mechanism underlying our result is completely different from theirs (changes in the bargaining position of a ship through varying its outside option value). Specifically, the dampening force in our model is directly related to the notion of congestion in shipping markets, which has not yet been studied before. Furthermore, we follow the suggestion in Brancaccio et al. (2020) by endogenizing oil prices in a general equilibrium framework involving

exporters, ships, and a final good producer. We then show quantitatively that stabilization in the oil trade and in the final good production is achieved at the cost of exaggerated oil price volatility, since the prices inherit the variations in tanker utilization between shipping markets. This finding offers a novel explanation for the excessive oil price volatility observed in reality (Chatziantoniou et al. (2021) and L. Liu et al. (2016)). In addition, by modeling the market switching behavior of tankers as they switch between oil shipping markets that are geographically remote, our findings also contribute to the large body of literature that studies the role of geography in generating transportation networks (Buchholz (2021), Fajgelbaum and Schaal (2020), and Frechette et al. (2019)).

Second, our paper links to the search and matching literature, as we build our theoretical model on a random search model in the spirit of the seminal work of Mortensen and Pissarides (1994), in which search frictions in the labor market were observed to prevent workers from perfectly matching up with firms. In order to account for the observed micro behaviors of oil tankers in the shipping data, while at the same time building the model in such a way that the macro implications of such behaviors are clear, we reconstruct the model of Pilossoph (2014) with two major innovations. First, the model permits a feedback loop such that shipping market congestion affects a ship's searching behavior within and between markets, while each ship's searching behavior in turn determines how congestion intensifies in a shipping market. Second, the model explicitly considers the three stages of the oil trade, namely, export, transportation, and final good production. This set-up makes it possible to study, for the first time in the literature, how variations in oil exports (that is, oil trade shocks) and transportation affect the oil trade, oil prices, and final good production.

Third, our findings contribute to the study of oil price determination, as in Bodenstein et al. (2011), Kilian (2009), and Kilian and Zhou (2018), and most recently, Bornstein et al. (2022) and Känzig (2021). Specifically, in his estimation of the dynamic effects of oil shocks (i.e., oil supply, aggregate demand, and precautionary demand shocks) on real oil prices, Kilian (2009) applies shipping

freight rates as a measure of global real economic activity that drives the demand for industrial commodities in global markets. Our model micro-founds spot freight rates (i.e., commissions) in the oil transportation industry, and rationalizes the application of freight rates as a measure of commodity demand.[4] Indeed, our model predicts that higher demand for oil translates to more commissions paid to oil tankers, even though such an increase gradually shrinks due to the emergence of shipping market congestion. In this regard, our paper also contributes to the literature on endogenous trade costs, such as Asturias (2020), Boddin and Stähler (2018), Brancaccio et al. (2020), Coşar and Demir (2018), and Wong (2018).

The remainder of the paper is organized as follows: Section 4.2 provides an overview of the oil transportation industry, describes the unique shipping data, and illustrates the three novel facts described above. Section 4.3 presents our novel model of the shipping markets, our parameterization strategy, and the corresponding estimation results. Section 4.4 discusses the model's significant implications, while Section 4.5 concludes. The appendices contain additional empirical evidence, proofs, counterfactual exercises, and estimation details.

4.2 Oil Transportation

4.2.1 Overview

The seaborne oil transportation industry comprises a limited number of oil tankers that each carry a homogeneous oil cargo in an individual trip. In this paper, we focus on long range (LR1 and LR2) product tankers.[5] They were originally designed to carry product oil (e.g., naphtha, gasoline, and jet fuel); however, on some occasions,

[4] In the process, we also discover that tanker utilization might serve as an appropriate measure of oil demand driven by fluctuations in the global business cycles. This measure has a potential advantage over freight rates since it could control for a situation where the fluctuations in freight rates are driven by the provision of shipping services which is external to global oil demand (e.g., the scrapping of emission-heavy oil tankers due to the enactment of IMO 2020). We leave its empirical application for future endeavors.

[5] For reference, LR1 and LR2 tankers are coated product tankers in the range of 55,000 to 125,000 deadweight tonnage (DWT).

they can opportunistically switch to the crude oil trade in order to take advantage of the freight earnings differentials between the two shipping markets.[6]

Trips are realized through individual contracts; shipowners have vessels for hire, and exporters have oil to transport. Ships carry at most one cargo at a time, as the exporter fills up the hired ship with the oil product. In this paper, we focus on spot contracts (also known as voyage charters), in which a ship is hired by an exporter to perform a specific trip.[7] During a voyage charter, the shipowner still manages the vessel, but the exporter selects the ports of origin and destination. The exporter pays for all the fuel that the tanker consumes, port charges, canal costs, and commissions that are specific to the voyage, while the shipowner covers the remaining shipping costs (e.g., the crew cost).[8]

4.2.2 Data

In order to study the micro behaviors of oil tankers, and to build the empirical foundation for the theoretical model that follows, we need a shipping dataset that is (1) longitudinal, (2) of high frequency, and more importantly, (3) tracks each oil tanker's shipping status. The longitudinal feature enables us to observe the outcome of each ship's job search, while the high frequency of the data reduces the time aggregation bias in the construction of contracting and arrival probabilities (Elsby et al. (2009), Shimer (2012)). In terms of each tanker's shipping status, we need information on whether the ship is (1) at high or low speed, (2) moving or in-port, (3) unloaded or loaded, and lastly, (4) if loaded, carrying crude or product oil (i.e., cargo type). The speed and proximity of a vessel to port are

[6]Although our sample is restricted by data availability and we cannot equally observe the micro behaviors of crude oil tankers (i.e., vessels dedicated to carrying crude oil), our framework can be easily extended once their data become available. Although crude oil tankers cannot ballast and switch between markets, their optimizing behaviors in equilibrium (dictated by either the shipowner or the shipping firm), unavoidably depend on the payoff of the job search and the corresponding outside option value.

[7]Voyage charters are the most common type of shipping contracts. Long-term contracts (time charters) do also exist, but they only represent a comparatively smaller share of contracts in the oil transportation industry. For instance, in the product oil market, around 26% (37%) of LR1 (LR2) tankers are employed through time charters.

[8]In practice, the shipping costs are generally not budgeted separately, but included in the negotiated spot rate.

critical for the identification of trips, while the load status and cargo type define the state to which the ship belongs.[9,10]

In order to construct such a shipping dataset, we take advantage of several data sources. The primary source is the AIS data of product tankers collected from 2017 to 2020. Originally designed to prevent ship collision, the AIS is an automatic tracking system installed on vessels. Its records provide real-time satellite and terrestrial data on a ship's static (e.g., ship identity, ship size), dynamic (e.g., ship position, speed, and time stamp), and voyage-related (e.g., draught, destination, and estimated time of arrival) features (D. Yang et al., 2019). With such data, we can identify a series of trips taken by oil tankers, for each of which it is known whether the tanker is transporting an oil cargo or not.

Informative as the AIS data are, they do not provide information on the cargo type of a product tanker, which, is of vital importance in our study as it is directly linked to the market switching behavior of these tankers. In order to pin down such information, we apply fixture data (i.e., shipping contracts) from Clarksons Shipping Intelligence Network. The data record a non-exhaustive list of transactions between shipowners and exporters during 2017-2020 and their respective features, such as, vessel name, laycan (laydays canceling, or L/C) dates, loading and discharge ports, shipowner, cargo type, and other characteristics.

Nevertheless, our fixture data only record the broker's own fixtures (i.e., deals closed by Clarksons) and hearsay in the market. Considering this, we also leverage the 2020 World Tanker Ports Map provided by Clarksons Shipping Intelligence Network. In the map, major tanker ports are categorized based on four designations, which are crude oil loading, crude oil discharge, product oil loading, and product oil discharge. Certain ports have only one designation. For example, ports like Ras Isa and Ash Shihr in Yemen can only load crude oil, whereas Caldera and

[9]Ideally, the shipping dataset should have a similar structure to the one in the Current Population Survey (Flood et al., 2020), with each ship's "employment" status classified as either loaded or unloaded, and its "occupation" being to carry either crude or product oil.

[10]Note that the idle state of ships (that is, temporarily shifting out of commercial operations and remaining idle in the market of last contract) cannot be distinguished from the unloaded state due to a lack of a unique identifier in the shipping data. See Appendix C.11 for a detailed discussion.

Guayacan ports in Chile can only discharge product oil. Given this information, we are able to pin down the cargo type of a product tanker through examining the designations of those ports where it loads or discharges oil cargo.[11]

While we can label a large proportion of loaded trips in our dataset by applying fixture data and port designation information, for those remaining loaded trips with unidentified cargo types, we apply the random forest (RF) ensemble learning technique (Breiman, 2001) so as to classify their cargo types. In a nutshell, this involves selecting a set of input variables that, in principle, have significant power to predict cargo types, and then applying bootstrap aggregation to build a large collection of uncorrelated classification trees and then average them.[12] With this methodology, we are able to classify any loaded trips with a cargo type unidentified from the previous steps, and complete the construction of the required shipping dataset.

Summary statistics. Our final dataset includes 757 oil tankers, which covers the entire fleet of product tankers operating between 2017 and 2020. We tally 24,950 loaded trips (with non-empty information), for each of which we know the cargo type, as well as other information such as trip duration, loading and discharge ports, etc. Among these trips, 9,075 (36.4%) are classified as crude oil trips, while the other 63.6% are classified as product oil trips. These loaded trips form the basis of our analysis: In them we can retrospectively observe the outcome of each ship's job search within and between markets, and calculate the corresponding transition probabilities. Lastly, there are a total of 83,079 ship-week observations in which ships are either partially or fully loaded, and another 53,331 ship-week

[11]Our method of using port information to classify cargo types is similar to the one used in Yan et al. (2020), where they apply information on the oil import and export attributes of a port to distinguish between the types of cargoes carried by oil tankers. We did not adopt their approach entirely for two main reasons: (1) the 2020 World Tanker Ports Map provides direct information on the type of oil cargoes that can be loaded or unloaded at a certain port, and (2) there are cases in which a tanker is partially loaded at an oil exporting port, and becomes loaded at an importing port.

[12]For instance, one input variable to consider is whether a loaded trip originates from the East or West of the Suez Canal. If the trip originates from the East of the Suez Canal, there is a high probability that the ship is carrying product oil, and vice versa.

observations in which ships are unloaded and choosing between staying in their current market or moving to the other.

4.2.3 Novel Facts

In this section, we document three novel facts about the oil transportation industry, namely, the presence of shipping market congestion (Section 4.2.3.1), the prevalence of market switching behavior by oil tankers (Section 4.2.3.2), and the propagation of oil trade shocks in the global economy through the transportation network (Section 4.2.3.3). The aim here is to highlight key aspects of oil shipping markets, which then provide the rationale for the mechanisms we will consider in the theoretical model.

4.2.3.1 Presence of Shipping Market Congestion

Shipping market congestion arises naturally from the imbalance between the number of unloaded ships searching for oil delivery jobs (the supply of shipping services) and exporters' capacity for hiring (the demand for shipping services).[13] From the perspective of ships, it implies that the presence of one additional ship searching within a market decreases the probability of others to make a successful match.[14]

One way of concisely illustrating the presence of congestion is through estimating the aggregate matching function in the oil transportation industry. This function, which has been used extensively in labor search and matching theory to study employment outcomes (e.g., Petrongolo and Pissarides (2001)), takes as inputs both the stock of unloaded and searching ships (s_t) and that of unmatched and searching exporters (v_t), and dictates the number of matches to be formed (λ_t):

$$\lambda_t = \Lambda(s_t, v_t), \tag{4.1}$$

[13]Shipping market congestion is closely linked to the presence of search frictions in the oil transportation industry. We document this in Appendix C.1 through the phenomenon of the under-utilization of empty ships in major, net oil exporting countries.

[14]From the exporters' perspective, the presence of shipping market congestion also implies that one additional exporter decreases the probability of a successful match for other exporters. Technically, we could illustrate this by applying a similar estimation framework to that in Section 4.2.3.1, i.e., calculating the average contracting probability for exporters, $q_t \equiv \Lambda(s_t, v_t)/v_t = \psi s_t^{\alpha} v_t^{-\alpha}$, and estimating its empirical relationship with the number of unmatched and searching exporters, v_t. This cannot be implemented from our dataset, however, since neither q_t nor v_t is observed.

where the function $\Lambda(.)$ is assumed to be increasing, concave, and linearly homogeneous in both of its arguments.[15,16]

We further consider a special case of the aggregate matching function. In particular, as in the vast majority of the search and matching literature (e.g., Elsby et al. (2009) and Petrongolo and Pissarides (2001)), we assume that the matching function is of the Cobb-Douglas form with constant returns to scale:

$$\Lambda(s_t, v_t) = \psi s_t^{\alpha} v_t^{1-\alpha}, \tag{4.2}$$

where ψ is a parameter that captures the matching efficiency, while α and $1-\alpha$ are the elasticities of matches with respect to searching ships and exporters. Now, given this matching function (4.2), we define the average contracting probability for ships (f_t) as:

$$f_t \equiv \frac{\Lambda(s_t, v_t)}{s_t} = \psi s_t^{\alpha-1} v_t^{1-\alpha}, \tag{4.3}$$

which is the probability that an average ship enters into a contract with an exporter to deliver an oil cargo at time t. Equation (4.3) implies that, based on the number of (unobserved) searching exporters (or, equivalently, the number of oil cargoes to be delivered), the elasticity parameter α essentially determines how the shipping market congestion intensifies when more ships decide to search for oil delivery jobs, i.e., how the average contracting probability changes when there is an increasing number of searching ships, holding all else constant. For instance, if α is small and lower than 1, it implies that increases in s_t will decrease f_t, generating congestion.

[15]That the matching function only depends on the number of searching agents on either side of the market implies that they are homogeneous in terms of their matching probabilities. For instance, the number of matches in an oil shipping market depends only on the number of searching tankers and oil exporters. Other characteristics (e.g., a vessel's age) do not significantly affect the matching rate, even if they determine the payoff of a job search or the freight rate. This is distinct from the matching literature that studies the pairings between heterogeneous agents (e.g., matching games). For instance, in an application to the theory of industrial organization, Fox et al. (2018) show that upstream firms sort to downstream firms based on the product quality and capacity of each of the firms.

[16]The assumption of linear homogeneity has gained some support in the literature on matching function estimation in the shipping industry, as well. Using dry bulk shipping data and a non-parametric estimation strategy as in Matzkin (2003), without assuming constant returns to scale, Brancaccio et al. (2020) show that the implied degree of homogeneity of the matching function under the Poisson distributed exporters is roughly equal to one.

Hence, by taking the natural logarithm of both sides of Equation (4.3), we derive Equation (4.4), though which we can estimate α by a linear regression:

$$\begin{aligned} \log(f_t) &= \log(\psi) + (\alpha - 1)\log(s_t) + (1 - \alpha)\log(v_t) \\ &= \beta_0 + \iota_t + \beta_1 \log(s_t), \end{aligned} \tag{4.4}$$

where $\beta_0 + \iota_t = \log(\psi) + (1 - \alpha)\log(v_t)$ and $\beta_1 = \alpha - 1$. Nevertheless, due to the problem of reverse causality (e.g., the average contracting probability also affects the number of searching ships in a market), the regression may lead to biased estimates. Therefore, we need some instruments to tackle the issue of endogeneity.

For this purpose, we apply sea weather conditions at 15 major ports located near the Persian Gulf as instruments for the number of searching ships in the oil transportation industry.[17] In theory, these weather instruments are valid because they affect the departures and arrivals of ships (and hence their searching behaviors) without affecting the number of exporters. Furthermore, they are relevant in the sense that a large proportion of crude and product oil trips originate from the Persian Gulf (see Figure 4.1 for an illustration), so any changes in sea weather there should, in principle, have an impact on the number of searching ships. In what follows, the validity and relevance of the weather instruments are formally tested, and their results are presented in Table 4.1.

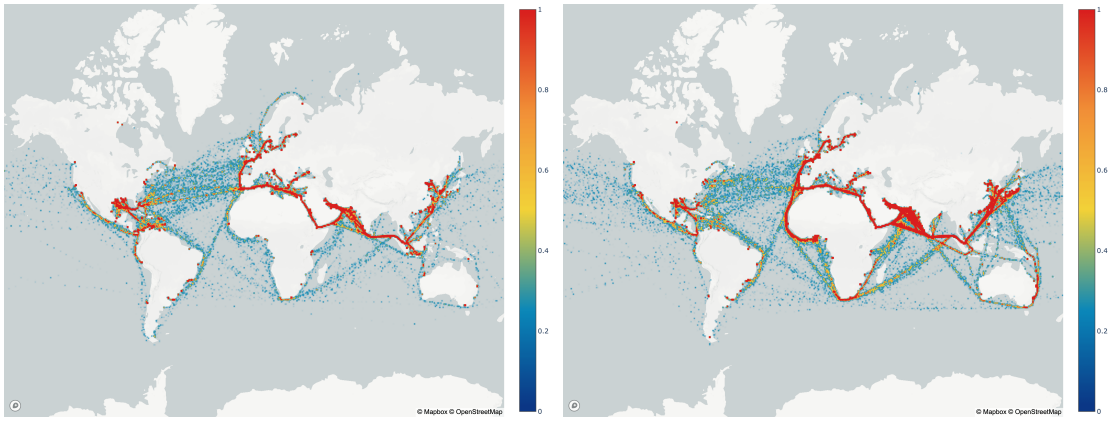
Table 4.1 shows the results of an instrument variables (IV) regression, in which we regress the log average contracting probability for ships on a constant, log average West Texas Intermediate (WTI) oil futures price (deflated according to the U.S. Consumer Price Index), and the number of unloaded and searching ships (which is normalized by the total number of ships), instrumented by the log average horizontal (E/W and N/S) speed of wind at 100m above the Earth and the average temperature of air at 2m above the surface of land, sea, or inland waters at 15 major ports near the Persian Gulf.[18] Note that the WTI oil futures price is included

[17] See Appendix C.2 on how we construct the weather instruments.

[18] The results are robust to alternative sea weather instruments, e.g., the average horizontal (E/W and N/S) speed of wind at 10m above the Earth, and the average horizontal (E/W and N/S) speed of the "neutral" wind at 10m above the Earth (the neutral wind is calculated from the surface stress and the corresponding roughness length by assuming that the air is neutrally stratified).

Figure 4.1: Geographic concentration of oil trade

(a) Crude oil trade **(b)** Product oil trade



Notes. The plots depict the geographic concentration of crude and product oil trips during the sampling period of 2017W5-2020W48. The trips are those with the cargo type identified using fixture data and port designation information. The geographic concentration of the oil trade is differentiated by colors in a spectrum, with red and blue representing high and low levels of concentration, respectively. The plots are created in Mapbox OpenStreetMap.

to control for the demand side effect on the average contracting probability for ships, and that the limited information maximum likelihood (LIML) is applied, as it is more robust to the weak instruments problem than 2SLS (Stock & Yogo, 2005). The first column summarizes the first-stage results: Even though the wind speed variables are not statistically significant, the F-statistic of the three weather instruments stands at 6.97, which exceeds the Stock-Yogo weak identification test critical value of 6.46 at the 5% level of significance and the 10% desired maximum size, indicating their joint relevance for the number of searching ships.

The estimated elasticity of the average contracting probability for ships with respect to the number of unloaded and searching ships ($\hat{\beta}_1$), as shown in the second column of Table 4.1, is equal to -0.614 and statistically significant. This implies that, for a one percent increase in the number of searching ships, the average contracting probability for all ships decreases by 0.6 percent, holding all else constant. This significantly negative elasticity clearly indicates the presence of shipping market congestion in the oil transportation industry. Lastly, the Sargan test of over-identification has a test statistic of 0.97 with a p-value equal to 0.62, meaning that

the weather instruments are valid and uncorrelated with the error term.[19]

Table 4.1: Estimation of the aggregate matching function

	OLS	**LIML**
	$\log(s_t)$	$\log(f_t)$
$\log(s_t)$		-0.614
		(0.284)
$\log(wind_{E/W,100m})$	-0.027	
	(0.021)	
$\log(wind_{N/S,100m})$	0.010	
	(0.020)	
$\log(temperature_{2m})$	1.079	
	(0.259)	
$\log(WTI_t)$	-0.037	0.096
	(0.016)	(0.023)
Observations	200	200
R^2	0.1238	0.4225
$F(3, 195)$	6.970	
Sargan statistic ($\chi^2(2)$)		0.966

Notes. LIML is applied to estimate the IV regression. As the test statistic of the Pagan-Hall test of heteroskedasticity stands at 7.393 with a p-value of 0.12, and that of the Cumby-Huizinga test of serial correlation in errors equals 0.943 with a p-value of 0.33, we apply homoskedastic standard errors in the IV regression. Data on the average contracting probability for ships and the number of searching ships are calculated using the shipping data in 2017W5-2020W48. Data on the wind speed and air temperature are obtained from the ERA5 dataset. Data on the WTI oil futures price are obtained from U.S. Energy Information Administration. Standard errors are reported in parentheses.

[19]In Appendix C.3, we follow Mercan et al. (2021) and present additional evidence using a vector autoregression (VAR) model in order to study the response of the average contracting probability for ships in the oil transportation industry to an exogenous increase in the number of unloaded and searching ships (i.e., a tanker under-utilization shock). The results in Figure C.2 clearly illustrate the presence of shipping market congestion.

4.2.3.2 Prevalence of Market Switching Behavior by Oil Tankers

Applying the shipping data, we measure the prevalence of market switching behavior by oil tankers through their corresponding market switching probability, which is defined as the fraction of ships that find oil cargoes to deliver which differ from those specified in their previous contracts. As seen in Table 4.2, in the crude oil market, this probability stands at an average of 0.37, indicating that on average more than one third of unloaded ships in the crude oil market that find oil to deliver are actually hired by product oil exporters. In the product oil market, the average market switching probability is lower, at 0.21, implying that roughly one out of every five unloaded ships in the product oil market that are freshly hired will need to carry crude oil on their next trip.

Table 4.2: Summary statistics of the market switching probability

	Mean	**S.D.**	**Median**	**Min**	**Max**
Market switching probability					
Crude oil → product oil	0.365	0.079	0.358	0.177	0.539
Product oil → crude oil	0.211	0.046	0.211	0.076	0.323

Notes. The market switching probability is defined as the fraction of ships that find oil cargoes to deliver that differ from those specified in their previous contracts. The summary statistics are calculated using the shipping data in 2017W5-2020W48.

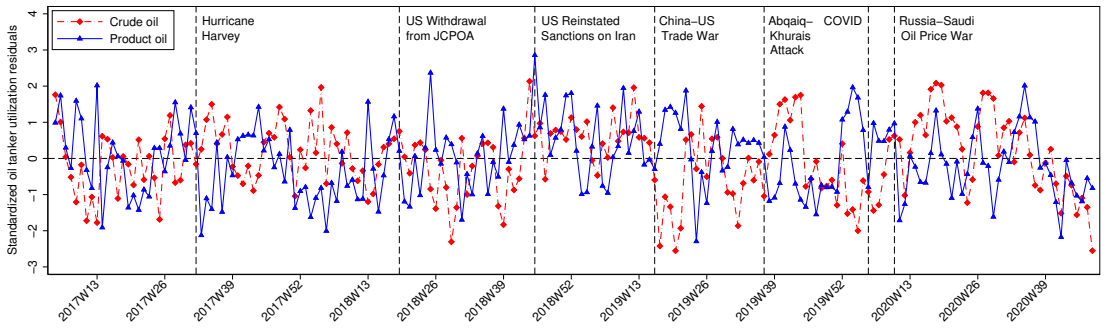
By comparing the two market switching probabilities, we can clearly see how frequently market switching takes place in the oil transportation industry, especially for crude oil tankers. Its prevalence, essentially, reflects the dispersion of oil trade shocks between the two shipping markets, as it is expected that most of the switches are triggered by differences in the market-specific payoffs induced by such shocks. This is true even though we cannot rule out the influence of other factors, e.g., seasonality in the oil trade, weather conditions, and maintenance or repair needs. In addition, the fact that oil tankers switch from the product to the crude oil trade much less often than the other way around is closely linked to the differences in the underlying structures of the two shipping markets. Notably, although the LR1 and LR2 tankers considered in our dataset are the main type of vessels used to

carry product oil, they are also the smallest carriers that can be used to serve the crude oil trade. In fact, due to the more homogeneous nature of seaborne crude oil transportation, larger vessels (e.g., very large crude carriers (VLCCs)) are designed for dedicated cross-regional, long-distance crude oil trade. This often gives rise to an escalation of congestion in the crude oil shipping market, which inevitably leads to a fall in the payoff of searching for cargo. It is only when the payoff of the crude oil trade is significantly higher than that of the product oil trade that product tankers will consider switching to the crude oil market.

4.2.3.3 Propagation of Oil Trade Shocks in the Global Economy through a Transportation Network

For this final fact, we present time series evidence to illustrate the propagation of oil trade shocks in the global economy through the oil transportation network. In particular, we present in Figure 4.2 the time series variations of tanker utilization in both crude and product oil shipping markets from 2017W5-2020W48.

Figure 4.2: Fluctuations in tanker utilization and major oil trade events from 2017W5-2020W48



Notes. The plotted series are the standardized residuals of tanker utilization in the crude and product oil shipping markets after controlling for monthly and yearly fixed effects from 2017W5-2020W48. Tanker utilization in an oil shipping market is defined as the proportion of tankers that are loaded with oil cargoes in that market over the entire fleet of product tankers. For reference, the mean tanker utilization in the crude oil shipping market during the sampling period is 0.200, while that in the product oil shipping market is 0.432. The vertical dashed lines indicate several major historical oil trade events, namely, Hurricane Harvey in 2017W32, the U.S. withdrawal from the Joint Comprehensive Plan of Action (JCPOA) in 2018W19, the reinstatement of sanctions by the U.S. on Iran in 2018W45, the China-U.S. trade war in 2019W16, the Abqaiq–Khurais attack in 2019W37, the COVID-19 pandemic in 2020W5, and the Russia-Saudi oil price war in 2020W10.

As seen in Figure 4.2, after controlling for seasonality, there have been significant variations in tanker utilization over the past four years. Specifically, we find that

the fluctuations in tanker utilization moved in mostly opposite directions in the crude and product oil markets, meaning that a rise in tanker utilization in one market often implied a corresponding fall in the other. In theory, these shifts in tanker utilization may arise due to a number of reasons (e.g., changes in contracting, arrival, or market switching probability), but here we focus on the differences in the contracting probabilities for ships, as they directly dictate how the spare capacity of oil transportation (i.e., unloaded ships) is distributed between the two shipping markets. For instance, suppose there is a positive trade shock that hits the crude oil market. On impact, the demand for shipping services to carry crude oil rises, which then leads to more commissions and a higher contracting probability for ships. These will raise tanker utilization in the crude oil shipping market and provide incentives for unloaded ships in the other market to switch over, which subsequently results in a decline of tanker utilization in the product oil market, holding all else constant.

Over the course of last four years, there have been several global oil trade disruptions, for instance, Hurricane Harvey in 2017, the Abqaiq–Khurais attack in 2019, the ongoing China-U.S. trade war, and the Russia-Saudi oil price war in 2020.[20] Their impacts have been far-reaching, and have not exempted the oil transportation industry. Using the shipping dataset, we visualize the impacts of these major oil trade disruptions on tanker utilization in Figure 4.2 so as to illustrate the propagation of oil trade shocks in the global economy through the oil transportation network. To start with, Hurricane Harvey was a devastating Category 4 hurricane that heavily impacted oil and gas production in the Gulf of Mexico and inland Texas. Quantitatively, Hurricane Harvey closed ports all along the Texas coast, reduced U.S. refinery capacity by 11% and oil production in the U.S. Gulf of Mexico by 25%, and led to higher gas prices (Scheyder & Seba, 2017). The fall in refinery capacity resulted in a corresponding decline in the demand for product oil transportation services, leading to a plummet in tanker utilization

[20]These major oil trade disruptions are considered exogenous to the oil transportation industry as they arise from either natural or political reasons. In Section C.13.1, we also study the newly enacted shipping regulations on vessels' sulfur emissions (i.e., IMO 2020) – which, in contrast, originates from the oil transportation industry – and its impacts on the oil trade, oil prices, and final good production.

in that market. Conversely, with the crude oil trade less affected, there was a simultaneous increase in tanker utilization in that market, as unloaded tankers switched over from the product oil market.

The trends in tanker utilization followed a similar pattern after the Abqaiq-Khurais attack in September 2019 that damaged Saudi Arabian oil production and provoked concerns over oil supply, and thus an increase in the demand for crude oil transportation (Hubbard et al., 2019).[21] At the same time, another shock also abetted the rise in tanker utilization in the crude oil market. On 25 September 2019, the U.S. added one of the largest shipping companies in China, COSCO, to its Specially Designated Nationals (SDN) Entity List, due to its transportation of Iranian crude oil after the re-imposition of sanctions by the U.S. against Iran. Often seen as another move by the U.S. against China in the trade war, in consequence, 20-25 VLCCs were idled and prohibited to carry crude oil (Gillis, 2020). This led to an increase in the demand for crude oil shipping services by product tankers, thus exaggerating the rise of tanker utilization in the crude oil market.

In 2020, as a result of high oil production in both Russia and Saudi Arabia despite massive demand destruction due to the COVID-19 pandemic, tanker utilization in the crude oil market rose as ships were used as floating storage for excess oil inventory. As shown in Figure 4.2, following its peak during the price war between the two countries, tanker utilization in the crude oil market dropped significantly, while at the same time, tanker utilization in the product oil market jumped. This can be explained as a result of skyrocketing freight earnings in the product oil market, leading tankers to switch markets. Due to massive demand destruction since the lockdowns, oil refineries reduced their operating rates, but not as swiftly as drops in demand. Consequently, there was a large surplus of product oil. Domestic inventories

[21]This is mostly related to the special role of Saudi Arabia in OPEC. It is by far the largest oil producer in OPEC, with enough capacity to function as a swing producer and balance the global market; it is often thought to be "OPEC's de facto leader," a title which is routinely used in the financial press, e.g., https://www.ft.com/content/1f84e444-9ceb-11e5-8ce1-f6219b685d74 (Accessed 13 July 2021).

quickly filled, and exporters used ships as floating storage, leading to excess demand for shipping services and high freight earnings in the product oil market.[22]

4.3 Baseline Model

In this section, we build a unique, multi-market, random search model of the shipping markets that accounts for the three novel facts of the oil transportation industry. The model is essential as it enables us to answer three important and related questions: (1) How do the oil transportation network and shipping market congestion adjust to exogenous oil trade shocks? (2) How does congestion affect the searching behaviors of ships within and between markets? (3) What are the implications of congestion for the oil trade, oil prices, and final good production? The model is built in such a way that the roles played by shipping market congestion, market switching behavior by oil tankers, and oil trade shocks in triggering macro effects can be easily disentangled from each other. This enables the execution of various counterfactual exercises. In addition, our model also provides a unified framework for policy-making and maritime event evaluation.

It should be noted that we develop the model in a setting with multiple commodity shipping markets for generalization purposes. The model is influenced by Pilossoph (2014) and Brancaccio et al. (2020). In what follows, we suppress the time indices until they become necessary for clarity.

4.3.1 Structure

The shipping industry is composed of N markets, each of which trades a homogeneous commodity. In each shipping market, unmatched exporters have commodity cargoes to deliver (which are subject to trade shocks), and search for available ships, while unloaded ships constantly search for commodity cargoes to deliver. The

[22] Another plausible explanation is related to the steep contango price structure (i.e., a situation where the futures price of oil is higher than its spot price) following the Russia-Saudi oil price war. In addition to product surplus due to insufficient cuts by refineries and full onshore inventories, the contango itself translated into more demand for ships as floating storage, because it is more expensive than onshore storage, and can only be used when contango is steep and/or freights are low.

matching process is random. If a match occurs, unloaded ships become loaded with cargo and start sailing, while the unmatched exporter pool is replenished by the entry of potential exporters. After loaded ships arrive at their destinations, the transported commodities are aggregated through a CES aggregator to produce a final consumption good. This completes the trade of each commodity as it transitions through export, transportation, and input to final good production.

In the baseline model, if unmatched to an exporter, an unloaded ship is allowed to switch markets and direct its search to some other market. If the ship chooses to switch markets, however, it must pay an additional cost in time spent traveling while unloaded. This state, known as *ballasting*, is distinct from remaining in the commodity market of its last contract and searching, because ships that are in the process of ballasting have no prospect of getting hired.[23] In addition, ballasting captures the time that ships must spend traveling to a new, geographically remote market. In the context of oil transportation, as shown in Figure 4.1, leaving the product for the crude oil market often implies a change of shipping route from the Persian Gulf to the Caribbean. When we estimate the model, the empirical counterparts that correspond to the stayers and movers are ships that remain in the same market as before and those that change operating markets, respectively.

4.3.1.1 Exports & Transportation

We start by introducing the process of export and transportation. There are $\mathscr{T}$ homogeneous ships in the economy. For each shipping market n, let t_n denote the total number of ships. The ships consist of loaded ships e_n and unloaded ships of two types – staying s_n and ballasting ships m_{nj}. Staying ships are those searching for contracts in commodity market n, while ballasting ships are those that were

[23]This is a modeling assumption, as there is no technological or other constraint that prevents a ballasting ship from searching en route. Nonetheless, there are economic disincentives that make matching unlikely, suggesting that it is a reasonable approximation. On the one hand, the geographic dispersion of oil shipping activities across regions of production, refining, and consumption implies that ships located farther away from a source are less competitive due to transportation costs. At the same time, exporters prefer ships in the same market as them because they value certainty and timeliness in shipping service, and ships in another, distant market cannot promise exact arrival times (Adland et al. (2016)).

last in market n, but are moving to search for jobs in some other market j. Hence, the total number of ships in market n is given by:

$$t_n = e_n + s_n + \sum_{j \neq n \in N} m_{nj}.$$

Search frictions. Commodity shipping markets are subject to standard search frictions. The probability of matching between ships and exporters in market n is determined by the market-specific matching function $\Lambda_n(s_n, v_n)$, where v_n represents the number of unmatched exporters (or equivalently, cargoes to be delivered). Staying ships can be hired instantaneously, and they remain unloaded simply because it takes time to find an exporter. Ballasting ships that are still in the process of traveling to some other market do not search en route. In the baseline model, we follow our results of matching function estimation in Section 4.2.3.1 and assume that Λ_n receives constant returns to scale:

$$\Lambda_n(s_n, v_n) = \psi_n s_n^{\alpha} v_n^{1-\alpha}, \tag{4.5}$$

where ψ_n is the market-specific matching efficiency and α is the elasticity of matches with respect to the number of staying ships. Letting $\theta_n \equiv v_n/s_n$ denote shipping market tightness, the probability that ships contract with exporters to deliver cargoes in market n is given by $f_n(\theta_n) = \Lambda_n(s_n, v_n)/s_n$, while the probability that exporters find ships to deliver commodities in market n is given by $q_n(\theta_n) = \Lambda_n(s_n, v_n)/v_n$. These two transition probabilities satisfy the relationship $f_n(\theta_n) = q_n(\theta_n)\theta_n$.

Market-specific trade shocks. In each shipping market n, the cargo delivered by each ship r_n is assumed to follow an autoregressive process so that:

$$\log(r'_n) - \log(\bar{r}_n) = \phi_n(\log(r_n) - \log(\bar{r}_n)) + \xi_n \nu'_n, \tag{4.6}$$

where $\bar{r}_n$ refers to the mean level of cargo delivered by each ship, ϕ_n to the autocorrelation in the de-trended natural logarithm of cargo delivered by each ship, $\nu'_n \sim N(0,1)$ to the exogenous, market-specific trade shocks, and ξ_n controls the variance of the series. We also allow for correlations in the trade shocks across

markets. Furthermore, we assume that cargoes can only be delivered by ships; each ship can carry (at most) one cargo, and each unmatched exporter can have (at most) one cargo for delivery.

Idiosyncratic taste shocks. In addition to trade shocks, following Artuç et al. (2010), Kennan and Walker (2011), and Pilossoph (2014), all ships draw a vector $\boldsymbol{\epsilon}$ of idiosyncratic taste shocks, $\epsilon_n \sim Gumbel(-\rho\gamma, \rho)$, every period.[24] The shocks are distributed independently and identically over time and across markets.[25] We interpret these taste shocks as anything that might keep ships in a market, and that is unrelated to commissions or the ease of contracting with an exporter. For example, a ship might not be able to find a cargo to deliver in market n, but continue to stay and search there because it is close to the ship's dry dock or bunkering. At the end of every period, and after realizing their tastes, unloaded ships are able to ballast to the market of their choice. This assumption ensures both that there are always some ships which will find it profitable to change markets (positive gross flows between markets), and that ship mobility is multi-directional (gross flows in excess of net flows), even in the absence of market-specific trade shocks – a prediction consistent with shipping data.

In addition, similar to Brancaccio et al. (2020), we assume that there are potential exporters in each market n that make a discrete choice on whether to export and are subject to independently and identically distributed taste shocks that follow $Gumbel(-\chi\gamma, \chi)$. Such a setting deviates from the free-entry condition that is usually assumed in the labor search and matching literature (i.e., jobs can be created instantaneously), but enriches our study as exporter entry now plays a

[24]The *Gumbel* distribution is also known as the Type I Extreme Value Distribution. Without loss of generality, for ships and potential exporters, we assume the means of their *Gumbel* distributions of taste shocks equal 0, which requires setting the shape parameters to $-\rho\gamma$ and $-\chi\gamma$ respectively, where $\gamma \approx 0.577$ is the Euler's constant.

[25]In an alternative setting, we could have modeled taste shocks correlated over time, which have the effect of significantly slowing the adjustment of ships in response to market-specific trade shocks. To differentiate this effect from that from shipping market congestion on the fluctuations of shipping market stocks and macro aggregates, we opt for the i.i.d. setting. In addition, autocorrelation in taste shocks adds unnecessary complexity, since we would need to keep track of the distribution of taste shocks within each market as an additional state variable when solving the model.

more explicit role in affecting shipping market tightness and the searching behaviors of ships within and between markets. In addition, the stock of unmatched exporters v_n becomes a state variable and is subject to variations in both the primitives (e.g., exit probability for unmatched exporters, quantity of potential exporters) and other endogenous variables (e.g., ballasting probability for ships, commissions). Furthermore, considering that commodity markets are often extremely concentrated with only a few major exporters, the free-entry condition seems less plausible due to high entry costs.[26]

Timing. Time is discrete, and the shipping industry begins with an allocation of ships (loaded, stayers, and movers) and exporters across all commodity shipping markets, given realized cargoes to be delivered $\{r_n\}_{n=1}^N$. Loaded ships transport cargo for a negotiated commission. Unloaded ships, depending on whether they stay in the same market as before or choose to ballast to some other market, incur a per-period waiting cost c^w, or a per-period sailing cost c^s.[27] Subsequently, the arrival of loaded ships, the contracting of staying ships, and the absorption of ballasting ships into new markets take place. The completion of each delivery job in market n is stochastic; a loaded ship arrives at its destination in the current period with probability d_n. Similarly, the arrival of ballasting ships is assumed to be stochastic; a ballasting ship arrives at its preferred market in the current period with probability d_{nj}. For exporters, those that get matched have their cargoes delivered and collect their revenue. Exporters that do not get matched wait at port. Only after these events occur do ships and potential exporters realize their taste shocks for the next period. Unloaded ships make a ballasting decision,

[26]In our context, the global oil market is dominated by a few big players. The biggest and most important player is OPEC, the 13 member countries of which account for around 44% of the world's crude oil production (based on data from U.S. Energy Information Administration for 2016).

[27]Note that for oil transportation, the pecuniary cost associated with switching markets includes not only the per-period fuel cost, but also a one-off tank cleaning cost. This is because, by industrial practice, the ship's cargo tank needs to be cleaned before it loads another type of oil product in order to avoid the risk of cargo admixture contamination (see https://maritime-mutual.com/blog/2020/11/09/cpp-tankers-standard-tank-cleaning-procedures-and-cleanliness-standards/ (Accessed 6 October 2021) for details). For simplicity, we assume that such a cleaning cost is instead spread evenly over the course of ballasting.

while potential exporters decide whether to enter the shipping market, thereby replenishing the exporter pool in the following period. After the reallocations between markets have taken place, the process starts over again after a new draw of cargoes to be delivered $\{r'_n\}_{n=1}^N$ has been realized.

Ships. As described earlier, there are three distinct states of ships: loaded, unloaded and staying in the same market as before, and unloaded and ballasting to a new market. Let W_n, S_n, and M_{nj} represent their respective values. First of all, the value of a loaded ship i in market n (net of idiosyncratic taste shocks) is given by:

$$\begin{aligned} W_n(\mathbf{\Omega}) =& w_n(\mathbf{\Omega}) + (1-d_n)\beta \mathbb{E}_{\mathbf{\Omega}',\boldsymbol{\epsilon}'}(W_n(\mathbf{\Omega}') + \epsilon'_{n,i}) \\ &+ d_n \beta \mathbb{E}_{\mathbf{\Omega}',\boldsymbol{\epsilon}'}\Big\{ \max\Big[S_n(\mathbf{\Omega}') + \epsilon'_{n,i}, \max_{j\neq n \in N}(M_{nj}(\mathbf{\Omega}') + \epsilon'_{j,i})\Big]\Big\}, \end{aligned} \tag{4.7}$$

where β is the time discount rate, $\mathbf{\Omega} = \{e_n, s_n, m_{nj}, v_n, r_n, \forall n, j \in N\}$ represents the state of the commodity shipping industry, and $\epsilon'_{n,i}, \epsilon'_{j,i}$ represent ship i's taste draws for markets n and j respectively for the next period. The present value of being a loaded ship in market n is the commissions w_n, plus the continuation value. With probability $1-d_n$, the ship continues its shipment in market n. With probability d_n, the ship fulfills its contract, becomes unloaded, and is able to choose whether to stay in market n, or to ballast from market n to some other market j. In this setting, we do not allow loaded ships to default on their contracts and search for new delivery jobs, either through ballasting to a new market or through job-to-job transitions. Apart from practical reasons, job-to-job transitions matter only insofar as we load all net reallocations on the currently unloaded.[28] While this might bias our estimate of net movements by unloaded ships upwards, it does not lead us to overestimate the amount of gross movement by unloaded ships, since we calibrate these flows to match the measures of ballasting ships in the shipping data.

[28]Job-to-job transitions are practically impossible in the oil transportation industry. Apart from the fact that a loaded ship cannot take on another delivery job with oil cargo from the previous job undischarged, there are strong economic disincentives for job-to-job transitions due to both the legal costs associated with a default on the previous contract, and to the damage it does to the shipowner's relationship with oil exporters.

Secondly, the value of an unloaded ship i staying in market n (net of idiosyncratic taste shocks) can be written as:

$$\begin{aligned} S_n(\mathbf{\Omega}) = &-c^w + f_n(\theta_n)\beta\mathbb{E}_{\mathbf{\Omega}',\boldsymbol{\epsilon}'}(W_n(\mathbf{\Omega}') + \epsilon'_{n,i}) \\ &+ (1 - f_n(\theta_n))\beta\mathbb{E}_{\mathbf{\Omega}',\boldsymbol{\epsilon}'}\Big\{ \max\Big[S_n(\mathbf{\Omega}') + \epsilon'_{n,i}, \max_{j\neq n\in N}(M_{nj}(\mathbf{\Omega}') + \epsilon'_{j,i})\Big]\Big\}. \end{aligned} \tag{4.8}$$

A staying ship in market n incurs the per-period waiting cost c^w, and earns the expected future value associated with staying in the same market as before. With probability $f_n(\theta_n)$, the ship enters into a contract with an exporter in market n. With probability $1-f_n(\theta_n)$, the ship does not contract with an exporter and chooses either to stay in shipping market n or ballasting to market j.

In contrast to staying ships, the value of an unloaded ship i ballasting from market n to j (net of idiosyncratic taste shocks) is given by:

$$\begin{aligned} M_{nj}(\mathbf{\Omega}) = &- c^s + (1 - d_{nj})\beta\mathbb{E}_{\mathbf{\Omega}',\boldsymbol{\epsilon}'}(M_{nj}(\mathbf{\Omega}') + \epsilon'_{j,i}) \\ &+ d_{nj}\beta\mathbb{E}_{\mathbf{\Omega}',\boldsymbol{\epsilon}'}\Big\{ \max\Big[S_j(\mathbf{\Omega}') + \epsilon'_{j,i}, \max_{k\neq j\in N}(M_{jk}(\mathbf{\Omega}') + \epsilon'_{k,i})\Big]\Big\}. \end{aligned} \tag{4.9}$$

A ship that travels from market n to j incurs the per-period sailing cost c^s, and earns the expected future value associated with ballasting. With probability $1-d_{nj}$, the ship continues its journey towards market j. With probability d_{nj}, the ship finishes its journey and arrives at market j. From there, it can choose either to stay or to ballast again to market k.

The unloaded ship's problem is thus, essentially, the choice between staying or ballasting. To simplify notation, define $\tilde{S}_{nj}$ as:

$$\tilde{S}_{nj} = \begin{cases} S_n, & \text{for } j = n; \\ M_{nj}, & \text{for } j \neq n \in N. \end{cases}$$

Given this simplification, the probability that an unloaded ship chooses to ballast from market n to j at time t is given by:

$$\pi_{nj}(\mathbf{\Omega}') = \Pr(\tilde{S}_{nj}(\mathbf{\Omega}') + \epsilon'_{j,i} > \tilde{S}_{nk}(\mathbf{\Omega}') + \epsilon'_{k,i}, \forall k \in N),$$

which can be re-written as:

$$\pi_{nj}(\mathbf{\Omega}') = \frac{1}{\sum_{k\in N}\exp\Big[(\tilde{S}_{nk}(\mathbf{\Omega}') - \tilde{S}_{nj}(\mathbf{\Omega}'))/\rho\Big]}. \tag{4.10}$$

The ballasting probabilities specified in Equation (4.10) are the standard logit probabilities that are widely used in the literature of labor reallocations (e.g., Caliendo et al. (2019) and Pilossoph (2014); they are also applied in Brancaccio et al. (2020) for dry bulk ships). The ballasting probabilities imply that, on average, unloaded ships ballast in response to differences in the market-specific payoffs. Lastly, to simplify the terms within the expectations in Equations (4.7), (4.8), and (4.9), we exploit the Type I Extreme Value Theory and integrate out future idiosyncratic taste shocks. The resulting value functions are presented in Appendix C.4.

Exporters. We now turn to the behaviors of exporters. Following closely the setup in Brancaccio et al. (2020), we begin with existing exporters and then consider entrants. First of all, the value of an exporter that is matched in market n can be written as:

$$J_n(\mathbf{\Omega}) = p_n r_n - w_n(\mathbf{\Omega}) + (1-d_n)\beta\mathbb{E}_{\mathbf{\Omega}'}(J_n(\mathbf{\Omega}')) + d_n\beta\mathbb{E}_{\mathbf{\Omega}'}(V_n(\mathbf{\Omega}')). \tag{4.11}$$

For every period when the delivery job is not yet fulfilled, the exporter earns revenue $p_n r_n$ and pays commission w_n to the contracted ship. With probability $1-d_n$, the match remains; with probability d_n, the contracted ship arrives at its destination, and the exporter becomes unmatched. The value of an exporter that remains unmatched in shipping market n is written as:

$$V_n(\mathbf{\Omega}) = (1-\delta)q_n(\theta_n)\beta\mathbb{E}_{\mathbf{\Omega}'}(J_n(\mathbf{\Omega}')) + (1-\delta)(1-q_n(\theta_n))\beta\mathbb{E}_{\mathbf{\Omega}'}(V_n(\mathbf{\Omega}')), \tag{4.12}$$

where δ represents the exit probability for unmatched exporters. In every period when the delivery job is not yet filled, the exporter keeps searching for an available ship and receives no payoff in the period. With probability δ, the exporter leaves the shipping market. Conditional on its survival, the exporter contracts with a ship to deliver the cargo with probability $q_n(\theta_n)$. With probability $1-q_n(\theta_n)$, the exporter continues searching. In addition, in every period in shipping market n, $\mathscr{E}_n$ potential exporters decide whether to export. They pay entry cost τ to enter the market. Each potential entrant then makes a discrete choice whether to engage in searching:

$$\mathbb{E}_{\mathbf{\Omega}',\boldsymbol{\epsilon}^{e\prime}}\Big[\max\Big(V_n(\mathbf{\Omega}') - \tau + \epsilon_n^{e\prime}, \epsilon_0^{e\prime}\Big)\Big],$$

where $\epsilon_n^{e\prime}, \epsilon_0^{e\prime}$ represent the potential entrant's taste draws for market n and not exporting, respectively, for the next period. We normalize the payoff of not exporting to zero for simplicity. Given that taste shocks $\boldsymbol{\epsilon}^{\boldsymbol{e}}$ are independently and identically distributed from $Gumbel(-\chi\gamma, \chi)$, we can write the choice probabilities underlying a potential exporter's behavior as:

$$\pi_n^e(\boldsymbol{\Omega}') = \frac{\exp[(V_n(\boldsymbol{\Omega}') - \tau)/\chi]}{1 + \exp[(V_n(\boldsymbol{\Omega}') - \tau)/\chi]}, \tag{4.13}$$

$$\pi_0^e(\boldsymbol{\Omega}') = \frac{1}{1 + \exp[(V_n(\boldsymbol{\Omega}') - \tau)/\chi]}. \tag{4.14}$$

Intuitively, when the market realizes a favorable trade shock, it leads to higher payoffs for exporters, and incentivizes potential entrants to enter shipping market n.

Commissions. When a ship and an exporter enter into a contract and negotiate commissions, they take into account each party's outside option. The outside option of the exporter is to remain unmatched and wait for another ship, while the outside option of the ship is to stay unloaded in the current market. The surplus of the match over the parties' outside options is split via the price-setting mechanism. The commission w_n paid to the ship delivering a cargo of valuation $p_n r_n$ in market n is determined by generalized Nash bargaining, with $\eta \in (0, 1)$ denoting the ship's bargaining power. Given the surplus sharing rule, i.e., $\eta(J_n(\boldsymbol{\Omega}) - V_n(\boldsymbol{\Omega})) = (1-\eta)(W_n(\boldsymbol{\Omega}) - S_n(\boldsymbol{\Omega}))$, we solve for the equilibrium commission w_n in the following lemma:

Lemma 1. *The agreed upon commission between a ship and an exporter in shipping market n is:*

$$\begin{aligned} w_n(\boldsymbol{\Omega}) = \eta p_n r_n - \eta\Big[(1 - \frac{\theta_n}{1-\delta})V_n(\theta_n) - (1-\theta_n)\beta\mathbb{E}_{\boldsymbol{\Omega}'}(V_n(\boldsymbol{\Omega}'))\Big] \\ - (1-\eta)\Big\{c^w + (1 - d_n - f_n(\theta_n))\beta\rho\mathbb{E}_{\boldsymbol{\Omega}'}\Big[\log\Big(1 - \sum_{j\neq n\in N}\pi_{nj}(\boldsymbol{\Omega}')\Big)\Big]\Big\}, \end{aligned} \tag{4.15}$$

where $\pi_{nj}(\boldsymbol{\Omega}')$ is the ship's ballasting probability from shipping market n to j.

Proof. See Appendix C.5. □

Commission negotiations resemble the Nash wage equation in the traditional search and matching framework with two innovations. First, commissions are affected by the exporter's outside option explicitly: Without the free-entry condition that is usually assumed in the search and matching literature, a higher outside option value V_n directly reduces the attractiveness of searching and hence lowers commissions, provided that the shipping market is less tight (i.e., θ_n is small, meaning that there are fewer cargoes to be delivered than unloaded ships in the market). Second, commissions include an extra term, which accounts for the fact that loaded ships are not allowed to ballast. As such, they must be compensated for the value of ballasting; the higher the ballasting probability π_{nj} in the next period, the higher the commission negotiated. This term significantly enriches the study of variations in shipping commissions, as it captures the attractiveness of both the current and all alternative shipping markets, which themselves display numerous features. For instance, markets that are unappealing to ships because there are few exporters relative to ships and the probability of ballasting afterwards is high, command higher commissions. The same holds for markets in which ships travel for longer periods to deliver cargoes, and those with low value exporters or low contracting probabilities.

Shipping Market Stocks and Flows. The stock of loaded ships in market n at period $t+1$ is given by:

$$e'_n = (1 - d_n)e_n + f_n(\theta_n)s_n. \tag{4.16}$$

That is, a fraction $1 - d_n$ of loaded ships in market n from the last period remain loaded and traveling. The ships staying in market n comprise the inflow that contract with exporters to deliver commodity cargoes, $f_n(\theta_n)s_n$.

For unloaded ships, the stock of staying ships in market n at period $t+1$ is:

$$s'_n = \pi_{nn}(\mathbf{\Omega}')[(1 - f_n(\theta_n))s_n + d_n e_n + \sum_{j \neq n \in N} d_{jn} m_{jn}]. \tag{4.17}$$

A fraction $\pi_{nn}(\mathbf{\Omega}')(1 - f_n(\theta_n))$ of staying ships in market n from last period do not find cargoes to deliver and choose to remain stayers. The inflow consists of loaded

ships from last period that fulfill their contracts and choose to remain in market n, $\pi_{nn}(\mathbf{\Omega}')d_n e_n$, plus ballasting ships from all other markets that just arrive at market n and choose to stay there instead of ballasting again, $\pi_{nn}(\mathbf{\Omega}')\sum_{j\neq n\in N} d_{jn}m_{jn}$.

The stock of ballasting ships in market n moving to j at period $t+1$ is given by:

$$m'_{nj} = (1-d_{nj})m_{nj} + \pi_{nj}(\mathbf{\Omega}')[(1-f_n(\theta_n))s_n + d_n e_n + \sum_{k\neq n\in N} d_{kn}m_{kn}]. \quad (4.18)$$

A proportion $1-d_{nj}$ of ballasting ships in market n to j from last period are still en route, while $\pi_{nj}(\mathbf{\Omega}')(1-f_n(\theta_n))s_n$ of staying ships in market n are unable to contract with an exporter and choose to ballast towards market j. A fraction $\pi_{nj}(\mathbf{\Omega}')d_n$ of loaded ships in market n finish their delivery jobs and choose to ballast to j. Finally, some ballasting ships from market k that just arrive at n choose to turn their search effort to market j, rather than remaining stayers in market n.

Lastly, the stock of unmatched exporters (or equivalently, cargoes to be delivered) in shipping market n is given by:

$$v'_n = (1-\delta)(1-q_n(\theta_n))v_n + \mathscr{E}_n\pi^e_n(\mathbf{\Omega}'), \quad (4.19)$$

where $\mathscr{E}_n$ represents the stock of potential exporters. This implies that a fraction $1-\delta$ of unmatched exporters survive the exogenous exit, and that the exporter pool is refilled by potential exporters that find it profitable to engage in shipping.

4.3.1.2 Final Good Production

After describing the process of exports and transportation, we discuss in this section the process of final good production. In a nutshell, each of the N shipping markets trades and transports a homogeneous commodity every period, and a final good producer utilizes the transported commodities as inputs for the production of a final good Y. In addition, it is assumed that shipowners spend all their earned commissions on the consumption of the final good.

Commodity demand. Specifically, the final good producer solves the profit maximization problem as follows:

$$\max_{\{y_n\}_{n\in N}} PY - \sum_{n\in N} p_n y_n \quad s.t., \quad Y = [\sum_{n\in N}(\mu_n)^{\frac{1}{\sigma}}(y_n)^{\frac{\sigma-1}{\sigma}}]^{\frac{\sigma}{\sigma-1}}, \quad (4.20)$$

where σ refers to the elasticity of substitution between commodity inputs, and μ_n stands for commodity n's share of production in the final consumption good, so that $\sum_{n\in N}\mu_n = 1$. In the baseline model, we assume that the CES demand shares are fixed at their statistical averages, and are to be estimated in the SMM procedures. If we let P denote the price of final consumption good and p_n denote the price of each commodity, it follows that the optimal demand for each commodity by the final good producer y_n can be written as:

$$y_n = \frac{Y\mu_n}{(p_n/P)^\sigma}, \tag{4.21}$$

where $P = [\sum_{n\in N}\mu_n(p_n)^{1-\sigma}]^{\frac{1}{1-\sigma}}$ is the price index. This implies that each commodity market faces a downward sloping demand curve.

Market clearing condition. In equilibrium, as exporters contract with ships to transport the traded commodity, the demand for each commodity by the final good producer is met by the cargo delivered by each ship r_n, times the number of loaded ships e_n, i.e., $\forall n \in N$,

$$y_n = r_n e_n. \tag{4.22}$$

Endogenous commodity price. Lastly, before describing the other equilibrium conditions, it is worth noting that commodity prices are endogenously determined in the baseline model. By combining the CES demands with the market clearing condition, we solve for the equilibrium commodity price in the following lemma:

Lemma 2. *The equilibrium price of commodity n, p_n, is given by:*

$$p_n = \Big[\sum_{j\in N}(\mu_j)^{\frac{1}{\sigma}}(r_j e_j)^{\frac{\sigma-1}{\sigma}}\Big]^{\frac{1}{\sigma-1}}(\frac{\mu_n}{r_n e_n})^{\frac{1}{\sigma}}.$$

Proof. Substitute the CES aggregator into Equation (4.21), and replace y_n with the commodity market clearing condition in Equation (4.22). □

The equilibrium commodity price p_n reflects rich dynamics that take place in shipping markets. On the surface, by inverting the CES demand, it implies that

each commodity's price is negatively related to its own trade, and positively related to the trade of other commodities.[29] In addition, by factoring in the market clearing conditions that directly link each commodity's trade to ship utilization in each shipping market, we show that the behaviors of ships across *all* shipping markets determine the evolution of the equilibrium price of commodity n. Simultaneously, the ships' behaviors are influenced by the imbalance between the demand for, and the supply of, shipping services in each market as a result of the dispersion of realized trade shocks and variations in tastes. Furthermore, the fundamentals embedded in the parameters, e.g., μ_n and σ, also affect prices both directly and indirectly through their impacts on the behaviors of ships.

Another, equally important point in the construction of these endogenous commodity prices is that the commodity markets are cleared by not only prices, but also by shipping market congestion. Without transportation, the randomness of exports would feed directly into the fluctuations of commodity inputs and then final good production. By adding an active transportation stage in which ships can dynamically adjust their searching behaviors in response to shocks and congestion, the force of randomness in shaping commodity trade and final good production is weakened.

4.3.2 Equilibrium

In treating the final consumption good as the numéraire of this economy and normalizing the total number of ships to one (i.e., $\mathscr{T} = 1$), an equilibrium in the shipping markets is characterized by an allocation of ships and exporters over markets $\{e_{n,t}, s_{n,t}, m_{nj,t}, v_{n,t}, \forall n, j \in N\}_{t=1}^{\infty}$; a set of value functions $\{W_{n,t}, S_{n,t}, M_{nj,t}, J_{n,t}, V_{n,t}, \forall n, j \in N\}_{t=1}^{\infty}$; ballasting probabilities $\{\pi_{nj,t}, \forall n, j \in N\}_{t=1}^{\infty}$; entry probabilities $\{\pi_{n,t}^{e}, \forall n \in N\}_{t=1}^{\infty}$; commissions $\{w_{n,t}, \forall n \in N\}_{t=1}^{\infty}$; and commodity prices $\{p_{n,t}, \forall n \in N\}_{t=1}^{\infty}$, such that for each period t, given initial allocations $\{e_{n,0}, s_{n,0}, m_{nj,0}, v_{n,0},$

[29] Inverting the CES demand function yields $p_n = [\sum_{k\in N}(\mu_k)^{\frac{1}{\sigma}}(y_k)^{\frac{\sigma-1}{\sigma}}]^{\frac{1}{\sigma-1}}(\frac{\mu_n}{y_n})^{\frac{1}{\sigma}}$. It is simple to validate that $\frac{\partial p_n}{\partial y_n} = -\frac{1}{\sigma}[\sum_{k\in N}(\mu_k)^{\frac{1}{\sigma}}(y_k)^{\frac{\sigma-1}{\sigma}}]^{\frac{1}{\sigma-1}-1}[\sum_{k\neq n\in N}(\mu_k)^{\frac{1}{\sigma}}(y_k)^{\frac{\sigma-1}{\sigma}}](\frac{\mu_n}{y_n})^{\frac{1}{\sigma}}\frac{1}{y_n} < 0$, and $\frac{\partial p_n}{\partial y_j}\Big|_{j\neq n\in N} = \frac{1}{\sigma}[\sum_{k\in N}(\mu_k)^{\frac{1}{\sigma}}(y_k)^{\frac{\sigma-1}{\sigma}}]^{\frac{1}{\sigma-1}-1}(\frac{\mu_j}{y_j})^{\frac{1}{\sigma}}(\frac{\mu_n}{y_n})^{\frac{1}{\sigma}} > 0$, $\forall\sigma > 0$.

$\forall n, j \in N\}$ and the evolution of cargoes to be delivered for each commodity $\{r_{n,t}, \forall n \in N\}_{t=1}^{\infty}$:

1. Loaded and unloaded ships behave optimally according to value functions $\{W_{n,t}, S_{n,t}, M_{nj,t}, \forall n, j \in N\}$, as given in Equations (4.7), (4.8), and (4.9);
2. Unloaded ships choose where to search so as to maximize payoff, so that the ballasting probabilities satisfy Equation (4.10);
3. Matched and unmatched exporters behave optimally according to value functions $\{J_{n,t}, V_{n,t}, \forall n \in N\}$, as given in Equations (4.11) and (4.12);
4. Potential exporters in each market optimally choose whether to engage in shipping, so that entry probabilities satisfy Equation (4.13);
5. Commissions evolve according to Equation (4.15);
6. The stock of loaded ships in each market evolves according to Equation (4.16);
7. The stock of stayers in each market evolves according to Equation (4.17);
8. The stock of movers in each market evolves according to Equation (4.18);
9. The stock of unmatched exporters in each market evolves according to Equation (4.19);
10. Each commodity market clears as in Equation (4.22).

In addition, in Appendix C.6, we show from Proposition 3 that, under some regularity conditions, a steady-state exists for such an equilibrium.

4.3.3 Parameterization

In this section we explain the parameter values in the baseline model. As a benchmark, we set up the model at weekly frequency and work with two shipping markets, crude (c) and product oil (p). We choose this setting in order to study the quantitative implications of shipping market congestion for the oil trade, oil prices, and final good production. Given this two-market specification, there are 24 parameters controlling the dynamic system, of which 16 are calibrated externally. We estimate the other eight parameters using SMM procedures for three reasons: (1) No existing literature infers their statistical values; (2) their parameterization is vitally important in governing shipping market stock and flows; and, most

importantly, (3) through SMM estimation, we tune the model to match empirical regularities in the shipping data, thus ensuring that the implications of our analysis are well-grounded. The parameter values are summarized in Table 4.3.

Table 4.3: Model parameters and assigned values

Symbol	Description	Value
Calibrated parameters		
β	Time discount rate	0.999
σ	Elasticity of substitution	2
$\{\phi_c, \phi_p\}$	Persistence parameter	$\{0.625, 0.679\}$
$\{\xi_c, \xi_p\}$	Standard deviation of trade shocks	$\{0.062, 0.038\}$
$corr(\nu_c, \nu_p)$	Correlation between trade shocks	-0.071
c^s	Per-period sailing cost	-0.008
c^w	Per-period waiting cost	-0.017
$\{d_c, d_p\}$	Arrival probability within market	$\{0.046, 0.037\}$
$\{\mathscr{E}_c, \mathscr{E}_p\}$	Stock of potential exporters	$\{1.030, 0.800\}$
δ	Exit probability for unmatched exporters	0.15
α	Elasticity of matching function	0.4
η	Bargaining power of the ship	0.4
Estimated parameters		
$\bar{r}$	Mean level of oil delivered by each ship	1.002
μ_c	CES demand share for crude oil	0.226
ψ_c	Matching efficiency in crude oil market	0.157
ψ_p	Matching efficiency in product oil market	0.089
ρ	Scale parameter of *Gumbel* distribution	0.318
d_{cp}	Arrival probability across markets	0.451
τ	Entry cost for potential exporters	7.00
χ	Scale parameter of *Gumbel* distribution	0.198

Notes. The estimated parameters are averages of the last 20 elements pooled over 20 MCMC chains, each of which has a length of 2,000.

4.3.3.1 Calibrated Parameters

The first half of Table 4.3 dictates the calibrated parameters. To begin with, we choose a weekly discount rate of $\beta = 0.999$, corresponding to an interest rate of 4%, p.a.

We set the elasticity of substitution between crude and product oil in the production of final consumption good σ to 2, recognizing that there is no consensus on this value in the literature, and that the production of product oil requires crude oil as input (Balistreri et al., 2010; Melek et al., 2017; Németh et al., 2011). We choose an intermediate value of 2 by taking a stand on the substitutability between crude and product oil in final good production.

Next, to calibrate the parameters governing the evolution of oil cargo delivered by each ship $\{r_c, r_p\}$, we fit the autoregressive process by matching it with data from our shipping dataset. Recall that the autoregressive process is given by: $\forall n \in \{c, p\}$,

$$\log(r'_n) - \log(\bar{r}_n) = \phi_n(\log(r_n) - \log(\bar{r}_n)) + \xi_n \nu'_n,$$

where $\bar{r}_n$ refers to the mean level of cargo delivered by each ship, and ϕ_n to the autocorrelation in the de-trended natural logarithm of oil cargo delivered by each ship, $\nu'_n \sim N(0, 1)$ to exogenous, market-specific trade shocks, while ξ_n controls the variance of the series. Based on the constructed dataset, by our calculation, the standard deviation of the de-trended natural logarithm of tanker utilization in the crude oil shipping market is 0.079, while in the product oil shipping market its value is 0.052. The weekly autocorrelations in the two time series are 0.625 and 0.679, respectively; thus, we set $\{\phi_c, \phi_p\}$ directly to these values. Assuming $\nu_n \sim N(0, 1)$ and solving for the implied ξ_n yields $\xi_c = 0.062$ and $\xi_p = 0.038$. We also set the correlation between ν_c and ν_p at -0.071 to match the correlation in the de-trended natural logarithm of tanker utilization in the crude and product oil shipping markets.

For c^s and c^w, we apply the estimates of shipping costs in Brancaccio et al. (2020), even though their focus is instead on dry bulk shipping, and Handymax (40,000–60,000 DWT) and Panamax (60,000–100,000 DWT) carriers are over-sampled in their data. As reported in their Table 5, sailing costs are fairly homogeneous, averaging 44,000 USD per week or 0.8% of the exporter's valuation of cargo. Port waiting costs are more heterogeneous and larger in magnitude, ranging from 90,000 USD to 290,000 USD per week, and averaging 2.6% of the exporter's

valuation of cargo.[30] Hence, we calibrate c^s at 0.8% of the average level of oil cargo delivered by each ship, and, taking into account that staying ships might not always choose to stay in-port due to the high costs (by the model's assumption, they can move freely while searching for cargoes to deliver in the current market), we set c^w as the average of the estimated sailing and port waiting costs, that is, at 1.7% of the average level of oil cargo delivered by each ship.

For the ship's arrival probability within each shipping market, we calibrate its value based on its counterpart in the shipping data. Assuming that arrivals are stochastic, we set the arrival probabilities $\{d_c, d_p\}$ at 0.046 and 0.037, respectively.[31]

Regarding the stock of potential exporters $\{\mathscr{E}_c, \mathscr{E}_p\}$, we set them equal to the total world production of crude and product oil, respectively, thus assuming that total production serves as an upper bound to the trade volume in each shipping market. From world production data for crude and product oil from 2002-2019, the average weekly production of crude oil is approximately 486,000 thousand barrels, while that for product oil is 377,000 thousand barrels.[32] In order to transform production totals into a number of potential cargoes, we first divide total production by the average ship size, taking into account that (1) the LR1 and LR2 tankers have an average capacity of 500 and 750 thousand barrels, respectively; and, (2) among the 757 tankers of which we have detailed shipping data, 375 of them belong to the LR1 category whereas the others are LR2 tankers (Hamilton, 2014).[33] Then, to normalize the stock of potential exporters, we divide them by the total number of tankers.

[30] As pointed out in Brancaccio et al. (2020), consistent with industry narratives, waiting at port is costly, both due to direct port and security fees, as well as the rapid depreciation of the ship's machinery and electronics and anti-fouling costs caused by the accumulation of microorganisms during immobility.

[31] Unlike the arrival probability within each shipping market, the arrival probability across markets for a ballasting ship cannot be calculated directly using the constructed shipping data; since the ballasting state cannot be differentiated from the staying unloaded state, we cannot determine when ballasting actually ends. This problem is elaborated in Section 4.3.3.2

[32] Source: JODI-Oil World Database. For product oil, we aggregate weekly production of naphtha, motor and aviation gasoline, kerosene, and gas/diesel oil.

[33] The Average Freight Rate Assessment (AFRA) classifies tanker vessels according to their DWT. The approximate capacity of a ship in barrels is determined by using an estimated 90% of the ship's DWT.

For the exit probability for unmatched exporters δ, we calibrate it at 0.15, implying that there is a 15% probability that an unmatched exporter leaves the shipping market.

Lastly, following the results of matching function estimation in Section 4.2.3.1 where β_1 is estimated at -0.614, by the relationship between β_1 and $\alpha - 1$ as in Equation (4.4), the elasticity of matches with respect to unloaded and searching ships α is estimated at approximately 0.4. In a search and matching framework like ours, assuming the Hosios (1990) condition holds and the ship's bargaining power equals the elasticity of matches with respect to unloaded and searching ships, i.e., $\eta = \alpha$, an estimate of 0.4 indicates that oil tankers have less bargaining power than exporters. This is in contrast with the estimate by Brancaccio et al. (2020) at 0.7 for dry bulk shipowners, but mostly in line with the estimate by Regli (2019) at 0.24 for VLCCs. Such results are consistent with industrial practice, as oil shipowners tend to lock in deals early when the market is booming, while dry bulk shipowners tend to do the opposite.[34]

4.3.3.2 Estimated Parameters

In this section, we introduce the remaining parameters and the SMM procedures for their estimation. We opt for SMM instead of the Generalized Method of Moments (GMM) because, in our model, it is difficult to derive analytically the values of selected model moments as functions of the underlying data and parameters. Furthermore, the SMM approach also helps us focus on the ability of the model to explain specific moments in the shipping data. The results are then easy to interpret, and establish possible avenues for improving the model's fit.

We first make two normalizations to the parameters, one to the mean level of oil cargo delivered by each ship, and the other to the arrival probability across

[34]To the best of our knowledge, there is no literature that estimates the bargaining coefficient of tankers in the oil transportation industry with the specific size trajectory of ours, and the evidence for other industries is mixed. The estimate in Brancaccio et al. (2020) depends on the exporter's valuation of the cargo, which is directly linked to a weighted average of the prices of iron ore, coal, grain, steel, and urea. On the other hand, the estimate in Regli (2019) builds on a spatial matching function estimation that accounts for unobserved matches, which are overlooked in Brancaccio et al. (2020).

markets. In the model, the CES demand shares $\{\mu_c, \mu_p\}$, along with the mean levels of oil cargo delivered by each ship $\{\bar{r}_c, \bar{r}_p\}$, jointly determine the average tanker utilization in the crude and product oil shipping markets. In the baseline, without leveraging data on oil prices, we set $\bar{r}_c = \bar{r}_p = \bar{r}$ and include μ_c and $\bar{r}$ in the SMM estimation, recognizing that trade shocks affect both the demand and supply of oil.[35] In terms of the arrival probabilities across markets, we set $d_{cp} = d_{pc}$ so as to attenuate the differences of their impacts on the ballasting behavior of ships. The normalizations of $\bar{r}$ and d_{cp}, together with the constraint that CES demand shares add up to 1, result in a simplification of the dynamic system, so that only eight free parameters are to be estimated using SMM, i.e., $\{\bar{r}, \mu_c, \psi_c, \psi_p, \rho, d_{cp}, \tau, \chi\}$.

We further note that the moments generated by the model are not theoretically equivalent to what we measure in the constructed shipping data, even though the model permits and keeps track of the evolution in state variables across shipping markets, i.e., $\{e_{n,t}, s_{n,t}, m_{nj,t}, \forall n, j \in N\}_{t=1}^{\infty}$. This issue arises mainly because the ballasting state is unobserved in the shipping data, until those ships find oil cargoes to deliver in a different market.[36] To address this issue, we convert the model moments to their empirical counterparts, which we can retrospectively observe in the shipping data. This involves (1) expanding the state space by keeping track of each tanker's operating market of last contract; (2) formulating the state transition matrix of the corresponding Markov chain; and, (3) computing the resulting allocations of ships across different states.[37]

With the converted model moments, we estimate the eight free parameters that govern the eight data moments listed in Table 4.4 using SMM. Specifically, the

[35] One way to rationalize these trade shocks is to consider them natural or political shocks, like the trade events described in Section 4.2.3.3.

[36] For instance, suppose that a ballasting ship last contracted to carry crude oil just finished its voyage to the product oil shipping market. Unfortunately, it does not find any oil cargo to deliver there and instead finds it profitable to ballast again back to the crude oil shipping market. At the time the ship arrives and contracts with a crude oil exporter, it transitions from ballasting to loaded from the model's perspective, while in the data it will only be counted as a staying ship that just found oil cargo to deliver in the crude oil shipping market. Similar problems arise when the ship transitions through multiple ballasting states (in opposite directions) and ends up becoming staying unloaded, since the interim ballasting states cannot be distinguished from the staying unloaded state in the shipping data.

[37] Details on the process of converting model moments to their empirical counterparts can be found in Appendix C.7.

market-specific matching efficiencies $\{\psi_c, \psi_p\}$ determine the contracting probabilities for ships, and affect the distribution of loaded ships across markets. Likewise, the entry cost τ and the variance of taste shocks for potential exporters χ dictate how the stock of unmatched exporters evolves and affects contracting probability in each market. In addition, the average oil cargo delivered by each ship $\bar{r}$ and the CES demand share μ_c also influence the distribution of loaded ships across markets. In the estimation, we thus target the contracting probability for ships and tanker utilization in each shipping market, and include shipping market shares. Lastly, since both the variance of taste shocks for ships ρ and their arrival probability across markets d_{cp} control the transition probabilities in the shipping industry, we exploit the longitudinal dimension of our shipping data and target the market switching probabilities for ships in the SMM estimation.

Appendix C.8 spells out the SMM estimation in detail. Specifically, we opt for the MCMC approach proposed by Chernozhukov and Hong (2003).[38] Our estimation results are presented in Table 4.4. All the moments match reasonably well.

4.4 Model Implications

In the following sections, we highlight the role that oil transportation plays in the propagation of oil trade shocks to the wider economy. Through impulse response functions and counterfactual exercises, we illustrate how the endogenous adjustment of the oil transportation network in response to exogenous trade shocks and the resulting congestion stabilize the fluctuations of oil trade and final good production at the cost of exaggerated volatility in oil prices. Specifically, we explain in detail the roles of shipping market congestion, market switching behavior by oil tankers, and oil trade shocks in giving rise to such results.[39]

[38]The main advantage of this approach is that it only requires function evaluations, and thus the discontinuous jumps that occur as a result of parameter search do not cause the same problems that would otherwise arise with a gradient-based, extremum estimator (e.g., the objective function is not smooth). However, it also has its drawback since it requires a very long Markov chain, and consequently a significant number of function evaluations, each requiring the model to be solved and simulated.

[39]We also consider in Appendix C.13 three natural experiments to analyze how the micro behaviors of ships and exporters respond to variations in shipping policy (e.g., shipping fuel

Table 4.4: Data vs. model moments from the SMM estimation

Moments	**Data**	**Model**
Contracting probability for ships		
Crude oil	0.074	0.075
Product oil	0.062	0.063
Tanker utilization		
Crude oil	0.200	0.196
Product oil	0.432	0.430
Shipping market share		
Crude oil	0.317	0.313
Product oil	0.683	0.687
Market switching probability		
Crude oil $\rightarrow$ product oil	0.365	0.373
Product oil $\rightarrow$ crude oil	0.211	0.211

Notes. The contracting probability for ships is defined as the probability that an average ship enters into a contract with an exporter in a shipping market. Tanker utilization in an oil shipping market is defined as the proportion of tankers that are loaded with oil cargoes in that market over the entire fleet of product tankers. The shipping market share is defined as the proportion of loaded oil tankers in a market over the total number of loaded oil tankers. The market switching probability is defined as the fraction of ships that just find oil cargoes to deliver that differ from those specified in their previous contracts. Data moments are the averages computed from the shipping data from 2017W5-2020W48, while model moments are the averages of 10,000 simulations computed using the estimated model.

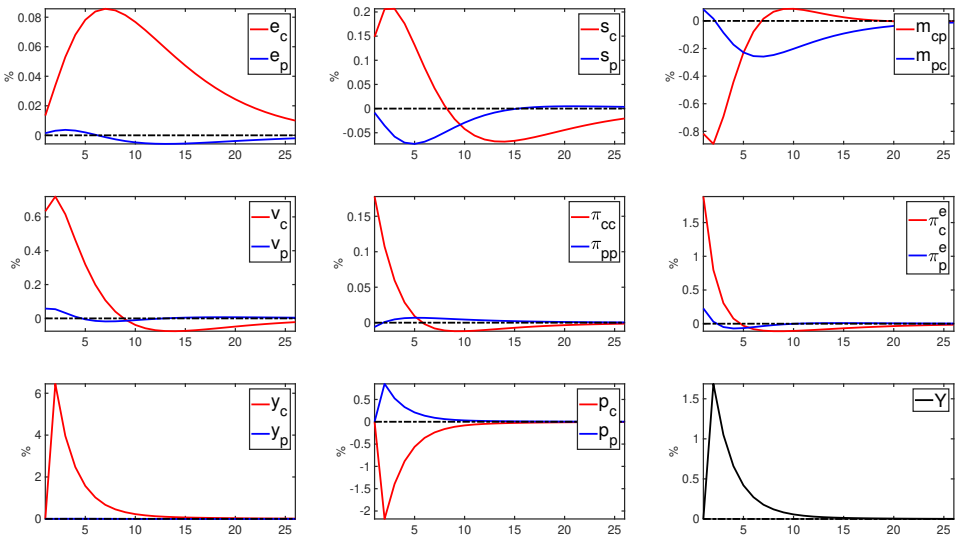
4.4.1 Impulse Response Functions

We use impulse responses functions (IRFs) to study how the oil transportation network endogenously adjusts to exogenous trade shocks. Figure 4.3 plots the IRFs of selected endogenous variables to an orthogonal, positive trade shock in the crude oil shipping market. Initially, the positive trade shock clearly favors a ship's job search in the crude oil shipping market, as e_c shoots up significantly. The rise in tanker utilization is intuitive. Due to the positive trade shock, the value of crude oil

regulations to cut sulfur emission), risk (e.g., the 2021 Suez Canal obstruction), and rationing schemes (e.g., the possibility of an unsanctioned Iran). Such experiments complement our findings on the role of oil transportation for the propagation of oil trade shocks in the wider economy, as they illustrate how "shocks" that are rooted in the oil transportation industry itself can trigger the endogenous adjustment of shipping markets and significant macroeconomic implications.

trade to exporters increases remarkably, and potential exporters find it profitable to engage in shipping, as shown by the initial spike of entry probability π_c^e. As a result, the stock of unmatched exporters v_c soars, giving rise to more demand for transportation services. Such excess demand then leads to greater tightness θ_c, more commissions w_c, and a higher contracting probability for ships f_c. For ships, this significantly enhances the value of staying in the crude oil market, while boosting the value of ballasting from the product oil market. The changes in value translate to a corresponding adjustment in staying probabilities, π_{cc} and π_{pp}, and as a result, the stock of ballasting ships in the crude oil market m_{cp} declines, while the stock of ballasting ships in the product oil market m_{pc} rises (though only marginally when compared to the size of the decline in m_{cp}). The impact of such initial adjustment on macro aggregates, coupled with the rise in tanker utilization is such that the trade in crude oil y_c increases dramatically, while the trade in product oil y_p hardly changes. The fluctuations in trade volume translate into inverse price adjustments, as the crude oil price p_c falls and the product oil price p_p rises. Lastly, as a result of the boost in the crude oil trade, final good production Y jumps significantly.

Figure 4.3: IRFs to a positive crude oil trade shock



Notes. The IRFs plot the dynamic weekly responses of selected endogenous variables to one standard deviation of an orthogonal trade shock in the crude oil shipping market ν_c. The IRFs are computed using a second order approximation of the baseline model, and are measured in percentage deviations from the steady-state.

Not long after the trade shock, the benefits accrued to ships searching in the crude oil market start to decline. As more ships flow into the market and the trade shock gradually fades away, tightness θ_c inevitably decreases; simultaneously, the amount of crude oil to be delivered v_c shrinks because fewer potential exporters decide to enter the market. The supply of shipping services thus becomes excessive due to growth in the stock of staying ships s_c as ballasting ships from the previous periods get absorbed. The resulting congestion then leads to declines in commissions and the contracting probability for ships. Consequently, ships that used to carry product oil may once again find it more profitable to continue searching in the product oil market, while those originally searching for crude oil are incentivized to switch markets. This pattern can be clearly seen from the third panel on top in Figure 4.3, as m_{cp} rises and m_{pc} falls after impact. The stock of staying ships, along with the staying probabilities, also adjust accordingly. Yet, we note that such a reversal of ballasting patterns does not trigger growth in tanker utilization in the product oil market, as not only does the substitution effect between crude and product oil weaken, but also the product oil shipping market's own congestion escalates.[40] This process continues until the mean-reverting behavior of the dynamic system takes hold, as the values of endogenous variables gradually return to their steady-state levels. Overall, a positive crude oil trade shock induces inverse movements in tanker utilization in the crude and product oil shipping markets, which is consistent with empirical evidence presented in Figure 4.2.

When the economy is hit by an orthogonal, positive trade shock in the product oil market, as shown by the sharp increase of e_p in Figure C.6, we can see that, on impact, searching in the product oil shipping market is clearly favored. However, the convolution of the substitution effect between crude and product oil and congestion

[40] On a side note, the substitution effect explains the initial increase in tanker utilization in the product oil market e_p as seen in Figure 4.3. As the crude oil trade y_c increases immediately following the positive trade shock, the price of product oil p_p will also rise. The price hike then encourages more product oil to be delivered v_p, raises tightness θ_p, and leads to higher tanker utilization e_p in the product oil market. In the baseline scenario, however, the substitution effect is second order when compared to the ship's ballasting motive. Therefore, the on-impact increase in tanker utilization in the product oil shipping market cannot be sustained due to a shrinking stock of searching ships s_p.

in the product oil market kicks in more quickly and significantly. Not only is the on-impact response of m_{cp} negative (indicating that, as soon as the trade shock hits, fewer ships decide to ballast to the product oil market and search there), but also the staying probability in the crude oil market π_{cc} increases immediately (indicating that more ships decide to search in the crude oil market). This increase in the supply of shipping services is met by an increasing demand, as more potential exporters are intrigued by the higher price p_c to enter the crude oil market and look for ships to deliver crude oil. Therefore, the on-impact increase in e_c is more dramatic than that in the previous case for e_p.[41] Nevertheless, despite the magnified response of e_c in the beginning, as the trade shock subsides, its growth will dive downward for precisely the same reasons as in the previous case: the substitution effect weakens, and the crude oil shipping market's own congestion amplifies. Therefore, on a longer time horizon, a positive product oil trade shock still induces inverse movement in tanker utilization in the crude and product oil shipping markets, which is also consistent with empirical evidence in Figure 4.2.

4.4.2 Counterfactual Exercises

As the previous section shows, shipping market congestion, market switching behavior by tankers, and oil trade shocks are interlinked with each other. The natural question to ask is what the implications of this intertwining are for the oil trade, oil prices, and final good production. In this section, we answer this question by performing three counterfactual exercises – full utilization, no ballasting, and zero trade variation – and comparing each of them with the baseline scenario.[42] They are designed to ensure that the roles played by shipping market congestion,

[41] Furthermore, if we compare the responses of tanker utilization between the two cases and dig deeper into the cause, the greater reaction of e_c in the case of a product oil trade shock can be largely attributed to the significantly larger substitution effect between crude and product oil in the process of final good production. As illustrated by shipping market shares in Table 4.4, the crude oil market attracts approximately 30% of loaded ships on average; therefore, the greater demand for crude oil triggered by the substitution effect will need to be accommodated by a proportionally larger increase of tanker utilization in the crude oil shipping market.

[42] In Appendices C.11 and C.12, we perform another two counterfactual exercises (the idling of ships and the free-entry of potential exporters), and compare each of them to the baseline scenario.

market switching behavior by tankers, and oil trade shocks in triggering macro responses can be easily disentangled from each other.

To run the counterfactual exercises, we need to establish the baseline first. We recover the trade shocks that hit the crude and product oil shipping markets from 2017W5 to 2020W48, and that are consistent with the observed evolution of tanker utilization in the shipping data, conditional on the estimated parameters in Table 4.3. Specifically, we take a first order approximation of the baseline model around its deterministic steady-state, which gives an approximate rule for how the model's endogenous variables evolve following exogenous trade shocks to r_n, the oil cargo delivered by each ship. Subsequently, combining the first order approximation with the time series of tanker utilization implies that solving for the underlying shocks boils down to solving a system of two non-linear equations with respect to the shocks in every period, $\nu_{c,t}$ and $\nu_{p,t}$. We assume that the model economy is at a steady-state in 2017W4, and then back out the time paths of market-specific trade shocks $\{\hat{\nu}_{c,t}, \hat{\nu}_{p,t}\}$ and other endogenous variables that are consistent with the observed evolution of tanker utilization in the shipping data.[43] Lastly, we use the implied oil trade, oil prices, and final good production, along with the observed tanker utilization, as our baseline for the counterfactuals.

4.4.2.1 Full Utilization

We start with a full utilization counterfactual in which the oil transportation industry is no longer "active" in the sense that all tankers are utilized to carry oil cargoes. This counterfactual shuts down the endogenous adjustment of the oil transportation network to exogenous trade shocks, and eliminates shipping market congestion. Accordingly, the oil trade can be defined as:

$$y_{n,t}^{F.E.} = \hat{r}_{n,t} \frac{\bar{e}_n}{\sum_{k \in N} \bar{e}_k},$$

where $\hat{r}_{n,t}$ stands for the data-consistent level of oil cargo delivered by each ship and $\bar{e}_n$ for the steady-state level of tanker utilization in shipping market n. By

[43] A complete description on how we recover the trade shocks is given in Appendix C.9.

Table 4.5: Baseline vs. counterfactual exercises

Scenario	e_c	e_p	y_c	y_p	p_c	p_p	Y
Baseline	0.200	0.432	0.224	0.487	0.843	1.058	0.704
	(0.016)	(0.023)	(12.387)	(15.071)	(27.352)	(12.481)	(12.924)
Counterfactual exercises							
Full utilization	0.313	0.687	0.328	0.753	0.433	0.997	0.989
	(0.000)	(0.000)	(19.798)	(24.084)	(0.847)	(0.828)	(22.873)
No ballasting	0.222	0.437	0.248	0.492	0.815	1.071	0.729
	(0.015)	(0.019)	(13.912)	(15.218)	(25.945)	(13.166)	(13.306)
Zero trade variation	0.200	0.432	0.201	0.433	0.842	1.059	0.627
	(0.016)	(0.023)	(0.016)	(0.023)	(0.031)	(0.014)	(0.026)

Notes. The above table shows the first and second moments of tanker utilization, oil trade, oil prices, and final good production for each of the four scenarios: baseline, full utilization, no ballasting, and zero trade variation. $\{e_c, e_p\}$ refer to tanker utilization in the crude and product oil shipping markets, $\{y_c, y_p\}$ to crude and product oil trade, $\{p_c, p_p\}$ to crude and product oil prices, and Y to final good production. The standard deviations are given in parentheses.

comparison, the oil trade in the baseline scenario is defined by $y^B_{n,t} = \hat{r}_{n,t}\hat{e}_{n,t}$. We essentially replace the observed tanker utilization $\hat{e}_{n,t}$ with the steady-state shipping market share $\bar{e}_n/(\sum_{k\in N}\bar{e}_k)$. We also calculate the corresponding oil prices and final good production following Equations (4.21) and (4.20).

Table 4.5 shows how the full utilization counterfactual compares to the baseline scenario in terms of macro aggregates. First of all, full utilization guarantees greater trade in both markets, as all ships are utilized to deliver oil; specifically, average crude and product oil trade increases by 46% and 55%, respectively. The average final good production, as described in Equation (4.20), is also larger because of greater oil inputs. According to the inverse relationship between the oil trade and price embedded in the downward sloping demand curve, increased trade generally implies lower prices. This holds true for both crude and product oil in the full utilization counterfactual.

Next, if we compare the standard deviations in Table 4.5, we can see that allowing for shipping market congestion is crucial, as it smooths the fluctuations in the oil trade and final good production caused by trade shocks. In fact, congestion stabilizes more than one third of the volatility in the oil trade and final good

production that would otherwise have occurred due to the impact of oil trade shocks. The stabilization is intuitive: Without congestion in the shipping markets, unloaded ships could find oil cargoes to deliver instantaneously, which implies that all the exogenous variations from oil trade shocks would translate to swings in the oil trade and final good production with no offsets from the oil transportation network.

Although shipping market congestion attenuates the fluctuations in the oil trade and final good production, it simultaneously exacerbates the volatility of oil prices, a pattern which can be clearly seen from their greater standard deviations in the baseline scenario from Table 4.5. This volatility in price can be attributed to two underlying forces, namely, the substitution effect inherent in the formation of oil prices, and the extra variations triggered by the endogenous adjustment of the oil transportation network to exogenous trade shocks. From Lemma 2, the equilibrium oil price in the baseline can be written as:

$$p_{n,t}^{B.} = \Big[\sum_{j \in N} (\mu_j)^{\frac{1}{\sigma}} (\hat{r}_{j,t} \hat{e}_{j,t})^{\frac{\sigma-1}{\sigma}} \Big]^{\frac{1}{\sigma-1}} (\frac{\mu_n}{\hat{r}_{n,t} \hat{e}_{n,t}})^{\frac{1}{\sigma}}, \tag{4.23}$$

while that in the full utilization counterfactual is given by:

$$p_{n,t}^{F.E.} = \Big[\sum_{j \in N} (\mu_j)^{\frac{1}{\sigma}} (\hat{r}_{j,t} \frac{\bar{e}_j}{\sum_{k \in N} \bar{e}_k})^{\frac{\sigma-1}{\sigma}} \Big]^{\frac{1}{\sigma-1}} \Big[\frac{\mu_n}{\hat{r}_{n,t} (\bar{e}_n / \sum_{k \in N} \bar{e}_k)} \Big]^{\frac{1}{\sigma}}, \tag{4.24}$$

where μ_n is the CES demand share for each oil input. The substitution effect arises naturally from the construction of CES demand, as each oil price is related to not only its own trade, but also that of the other oil product. Therefore, the volatility in both the crude and product oil trade adds to the fluctuations in each oil price, even though the trade itself is stabilized. Such fluctuations are further magnified by the endogenous adjustment of the ships' and exporters' responses to trade shocks. As seen in Equation (4.23), their behaviors affect the evolution of each oil price through $\hat{e}_{n,t}$, while their behaviors themselves are influenced by the dispersion of realized trade shocks across markets and other primitives of the model.

4.4.2.2 No Ballasting

Having illustrated the macro implications of shipping market congestion, we turn to the market switching behavior of oil tankers and its trade implications.

We start by hypothesizing a counterfactual in which oil tankers are no longer allowed to ballast between shipping markets. By doing so, we shut down the inter-market connection between crude and product oil transportation, leaving only final good production (which uses both as inputs) as the link between them. This implies that the value functions of ships must be rewritten to reflect their inability to ballast. Specifically, while the value of ballasting $M_{nj}(\mathbf{\Omega})$ no longer applies, the value of a loaded ship in market n becomes:

$$W_n(\mathbf{\Omega}) = w_n(\mathbf{\Omega}) + \beta \mathbb{E}_{\mathbf{\Omega}'}[(1 - d_n)W_n(\mathbf{\Omega}') + d_n S_n(\mathbf{\Omega}')],$$

as it can only transition into staying unloaded in its current market with probability d_n in the following period.[44] Likewise, the value of an unloaded ship in market n becomes:

$$S_n(\mathbf{\Omega}) = -c^w + \beta \mathbb{E}_{\mathbf{\Omega}'}[f_n(\theta_n)W_n(\mathbf{\Omega}') + (1 - f_n(\theta_n))S_n(\mathbf{\Omega}')],$$

as it can only become loaded in the current market with contracting probability $f_n(\theta_n)$ in the following period. By capping the size of each shipping market at the steady-state level of the shipping market share calculated in the baseline scenario, we can build an alternative, non-ballasting model using the same set of parameters as in Table 4.3. After applying a first order approximation, we solve for the implied tanker utilization, oil trade levels, oil prices, and final good production using the trade shocks which we recovered in the baseline scenario $\{\hat{\nu}_{c,t}, \hat{\nu}_{p,t}\}$. The means and standard deviations of these variables are summarized in Table 4.5.

What happens to the average level of the oil trade? First, we note that average tanker utilization in the crude oil shipping market in this counterfactual is significantly higher (by 11%) than that in the baseline. This happens because the average stock of staying ships in the crude oil market rises once the option

[44] In the no ballasting counterfactual, idiosyncratic taste shocks also no longer apply.

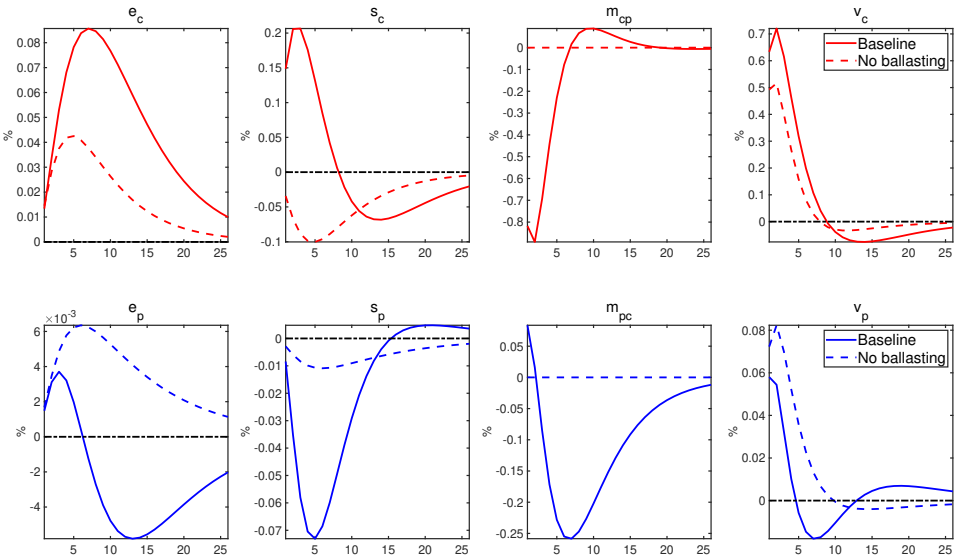
of ballasting is removed and they must stay and search for crude oil to deliver.[45] As for the product oil market, average tanker utilization increases only marginally because the outflow of ballasting ships that was possible in the baseline is minimal; thus, restricting ship movement does not change the stock of stayers in the product oil market significantly. As a result, the average level of the crude oil trade rises and the average level of the product oil trade remains relatively constant.

What happens to the volatility of the oil trade? Although higher averages in trade generally imply greater volatility, we still need to determine how the counterfactual model adjusts to trade shocks, and more importantly, how its dynamic responses to shocks differ from those associated with the baseline scenario. In Figure 4.4, we compare the IRFs of shipping market stocks to an orthogonal, positive crude oil trade shock in the baseline and no ballasting counterfactual. Note that the shock is comparable because the parameterization is identical, and the IRFs are measured in percentage deviations from the corresponding steady-state.

First of all, the on-impact response of tanker utilization in the no ballasting counterfactual is similar to that in the baseline: a positive trade shock incentivizes searching in the market where it hits. Nevertheless, the model dynamics that lead to a higher e_c are different between the two scenarios. In the baseline, the stock of both stayers and crude oil to be delivered jump on impact, with the rise in v_c greater than in s_c. This implies that the increasing demand for shipping services outweighs the rise in supply, producing a tighter shipping market, higher commissions, and greater ease in finding crude oil to deliver. In the no ballasting counterfactual, potential exporters are still interested in entering the crude oil market and taking advantage of the positive shock. Yet, since the market size is capped and there are no inflows of searching ships from the other market ($m_{pc} = 0$), a rise in e_c inevitably results in a corresponding fall in s_c. Consequently, as more exporters compete for a shrinking supply of shipping services in the crude oil market, the

[45] Recall that in the baseline model, the ballasting probability is always positive by construction since $\pi_{nj}(\mathbf{\Omega}') = 1/\sum_{k \in N} \exp\left[(\tilde{S}_{nk}(\mathbf{\Omega}') - \tilde{S}_{nj}(\mathbf{\Omega}'))/\rho\right] > 0, \forall j \neq n$. This implies that there are always gross flows between the crude and product oil shipping markets.

Figure 4.4: IRFs to a positive crude oil trade shock: Baseline vs. no ballasting



Notes. The IRFs plot the dynamic weekly responses of shipping market stocks (i.e., $\{e_{n,t}, s_{n,t}, m_{nj,t}, v_{n,t}, \forall n, j \in \{c, p\}\}$) to one standard deviation of an orthogonal trade shock in the crude oil shipping market ν_c for the baseline and no ballasting counterfactual, respectively. The shock ν_c is identical in both scenarios. The IRFs are computed using a second order approximation of each alternative model, and are measured in percentage deviations from the steady-state.

congestion lessens for ships but intensifies for exporters.[46] This leads to higher commissions and greater contracting probability for ships, thus encouraging tanker utilization in the crude oil shipping market.

Turning to the product oil shipping market, by isolating the impact of the substitution effect from the effect of ballasting, we can clearly see how influential the substitution effect is to the endogenous adjustment of shipping markets to exogenous shocks. Despite the magnitudes, the on-impact responses of e_p, s_p, and v_p are the same as those in the crude oil market, even though the underlying mechanisms that drive such responses are drastically different. In contrast to the direct boost to tanker utilization through the shock, the relative scarcity of product oil (as reflected in higher prices) encourages more potential exporters to enter the product oil market. Since there are no inflows of movers from the other side ($m_{cp} = 0$), the product oil market tightens as a result, leading to higher commissions and easing the

[46]The intensified shipping market congestion for exporters also explains why the initial response of v_c in the no ballasting counterfactual is relatively mild compared to that in the baseline.

finding of product oil to deliver. These benefits then encourage tanker utilization e_p and correspondingly deplete the stock of stayers s_p in the product oil market.

Next, we examine how allowing for the ship's ballasting behavior changes the initial tanker utilization responses in both the market where the shock hits and the other one. When the trade shock occurs, if allowed to ballast, ships are motivated to switch to the market where the shock hits and benefit from the higher payoff of searching there, as the substitution effect is only second order. As shown in Figure 4.4, the switches magnify the decline in s_p and at the same time, enlarge the pool of s_c. For the product oil market, this implies that the shortage of supply in transportation services becomes more severe, which consequently leads to a less sharp increase in tanker utilization than in the no ballasting case. In the crude oil market the opposite holds, as the shortage in shipping supply lessens due to the influx of ballasting ships. As a result, there is a more exaggerated increase in tanker utilization than in the case of no ballasting. This discussion also holds true in the case of a positive product oil trade shock. As shown in Figure C.7, when compared to the no ballasting counterfactual, the initial response of e_p in the baseline scenario is heightened, while that of e_c is muted.

The asymmetric tanker utilization responses in the two markets hold several implications in terms of volatility in the oil trade, and how these implications link to those by changing averages. First, whether the endogenous responses of the oil transportation network to exogenous trade shocks stabilize the market depends on which of the two counteracting forces associated with ballasting prevails. On the one hand, the greater increase in tanker utilization in the market in which a positive trade shock hits amplifies the increase in the oil trade and makes it more volatile. In this case, ballasting destabilizes the oil trade. On the other hand, when a positive trade shock hits the other market, the stabilization force of ballasting for the oil trade takes effect in the current market, because the increase in tanker utilization (and thus the oil trade) induced by the substitution effect is constrained by a shrinking stock of staying ships when more decide to ballast away.[47]

[47]In the event of a negative trade shock, the implications of ballasting for fluctuations in oil trade are the same. When a market is hit by an adverse trade shock, allowing ships to ballast

With this in mind, whether the stabilization force embedded in the endogenous responses of shipping markets can eventually reduce volatility in the oil trade depends crucially on the changes in average tanker utilization as a result of ballasting, since higher averages generally imply greater volatility. As shown in Table 4.5, the convolution of lower average tanker utilization and the stabilization force of ballasting dampens the fluctuations in the crude oil trade, as the standard deviation of the crude oil trade in the baseline is 11% smaller than that in the no ballasting counterfactual (12.39 vs. 13.91). Conversely, if we compare the standard deviations of the product oil trade (15.07 vs. 15.22), we can see that the stabilization is less effective, as the fluctuations in the product oil trade are hardly attenuated in the baseline.

Last but not least, how does the impact of ballasting on oil trade stabilization compare to that of shipping market congestion? In Table 4.5, by comparing the standard deviations of the crude and product oil trades respectively between the baseline, full utilization, and no ballasting scenarios, we conclude that it is *shipping market congestion* that contributes most significantly to oil trade stabilization: The crude and product oil markets see a reduction in volatility of 30% and 37%, respectively, when the tankers are simply allowed to become unloaded and search; further enabling them to ballast between markets only adds another 7.7% and 0.6% to crude and product oil trade stabilization, respectively.[48]

4.4.2.3 Zero Trade Variation

In this final counterfactual, we explore how trade shocks affect the oil trade, oil prices, and final good production. To do so, we consider a zero trade variation counterfactual in which there are no exogenous trade shocks, or equivalently, the oil cargo delivered by each ship is constant at its mean level. This counterfactual shuts

away will exacerbate the fall in the oil trade and add to its volatility; however, when the other market is hit by an adverse trade shock, ballasting ships will be attracted to the current market, which serves to mitigate the corresponding fall in the oil trade triggered by the substitution effect.

[48]Keeping the market size the same and assuming the process of final good production is identical, the only difference between the full utilization and no ballasting counterfactual exercises is whether shipping market congestion exists as a result of searching behaviors by ships and exporters.

down variations in the trade shocks while keeping the oil transportation network intact. Given this setting, the oil trade can be written as:

$$y_{n,t}^{Z.T.V.} = \bar{r}_n \hat{e}_{n,t},$$

where $\bar{r}_n$ represents the mean level of oil cargo delivered by each ship in market n, and $\hat{e}_{n,t}$ is the observed tanker utilization in the shipping data. We also calculate the corresponding oil prices and final good production. We can clearly see from Table 4.5 that, in this case, all the macro aggregates (including oil prices) fluctuate significantly less than in the other scenarios. This implies that it is the oil trade shocks that create most of the observed fluctuations in the oil trade, oil prices, and final good production; compared to these trade shocks, the corresponding endogenous adjustment of the oil transportation network and the resulting congestion effect only play a minimal role in the process.

4.5 Concluding Remarks

Our paper studies how the micro behaviors of oil tankers and their interactions with exporters in the oil transportation industry determine shipping market congestion, which is critical to the trade and prices of oil and final good production. We construct a novel shipping dataset that tracks the shipping status of 757 oil tankers from 2017-2020 and establish three novel facts about the industry, namely, the presence of shipping market congestion, the prevalence of market switching behavior by oil tankers, and the propagation of oil trade shocks in the economy through the transportation network. We also develop a novel search and matching model of oil transportation that not only accounts for these novel facts, but also embodies the three key stages of the oil trade: exports, transportation, and final good production.

With the estimated model, we study the congestion effect of oil transportation using three representative exercises. First, we show that shipping market congestion prevents the excessive flow of ships into a market that receives a favorable trade shock, such that the fluctuations in tanker utilization are attenuated. Second, we find that congestion plays a vital role in stabilizing the fluctuations in the oil

trade and final good production induced by oil trade shocks, as it attenuates over a third of the volatility that would otherwise have occurred. The stabilization is driven by shipping market congestion, as it dampens the dynamics of tanker utilization induced by oil trade shocks. We also find that such stabilization comes at a trade-off, as the volatility in oil prices increases by more than tenfold; this occurs because the pricing mechanism in our model inherits the dynamics in the oil shipping markets. Furthermore, we emphasize that, compared to exogenous trade shocks, the corresponding endogenous adjustment of the oil transportation network and the resulting congestion effects play only a minimal role in creating aggregate fluctuations.

5
Conclusion

The theme of this thesis rests at the intersection of individual heterogeneity, micro-level, search and matching behaviors of economic agents, and their corresponding, significant macroeconomic implications. All the three chapters evolve following this underlying theme.

Specifically, chapter 2 starts with an empirical observation: there have been lagging labor reallocations outside agriculture amid sustained, low agricultural productivity in the Philippines during the past 15 years. We then conduct an analysis of the labor adjustments in and out of agriculture and highlight that a variety of factors have influenced this process. We find that the widening of wage differentials with non-agricultural sectors, improvements in labor market efficiency, and better transport infrastructure are largely associated with growing outflows of labor from agriculture, whilst the lack of post-primary education and the presence of agricultural clusters hinder such outflows. Furthermore, in contrast to the traditional view that agricultural employment outflows are largely driven by productivity differences (as reflected by wage differentials), our results emphasize the roles of education as well as transport infrastructure in facilitating labor reallocations from agriculture to non-agriculture.

The rich information contained in the Philippines LFS data enables a variety of opportunities to be pursued; among all, one promising research topic is related to the

impacts of climate change and natural disasters (e.g., typhoons in the Philippines) on the within- and cross-region labor reallocations and how they may lead to both short-run (e.g., job losses and the associated disproportionately larger impacts on households with low financial buffer) and long-run (e.g., migration and the so-called "brain drain" effect which reduces the supply capacity in regions which were hit the hardest) implications.[1,2,3] This research can offer valuable insights for policy making so as to mitigate economic scarring and promote post-disaster recovery.

Chapters 3 and 4 use new granular data of oil tankers to study how the shipping market tightness (i.e., the imbalance between the demand for, and the supply of, shipping services) determines congestion in the oil transportation markets that is critical to the trade and prices of oil and production of a final good. We construct a unique shipping dataset that encompasses the entire fleet of product tankers during 2017-2020 using the state-of-the-art machine learning technique, and establish three novel facts: (1) the presence of shipping market congestion significantly reduces the contracting probability for oil tankers, (2) tankers switch between operating markets frequently, and (3) oil trade shocks propagate in the economy through a transportation network. We develop a novel search and matching model of oil transportation that centers on the formation of shipping market congestion; the model accounts for the novel facts and incorporates the interactions between oil exporters, tankers, and a final good producer in giving rise to congestion. We show with counterfactual exercises that congestion entails a trade-off as it attenuates over one third of the volatility in oil trade and production of a final good, while increasing oil price fluctuations by more than tenfold. We also show that the endogenous

[1]With an average of 20 typhoons a year, the Philippines is the third most disaster-prone country in the world. For instance, Typhoon Haiyan, known locally as "Yolanda", struck the Philippines on 8 November 2013, in what was reportedly the country's worst-ever natural disaster. More than 7,000 people lost their lives and six million workers lost their source of income. See https://www.bbc.co.uk/bitesize/guides/z9whg82/revision/4 (Accessed: 15 September 2021) for a detailed coverage.

[2]See Beine et al. (2001, 2008) and Bhagwati and Hamada (1974) for a discussion on the "brain drain" effect.

[3]Equally, the data may indicate otherwise as seasonal migration can be quite costly for households close to subsistence living and bear enormous risks (Bryan et al., 2014). If this were the case, an interesting question to ask could be what the key withholding forces are that keep households prone to natural disasters from migrating to safer regions.

adjustment of the oil transportation network and the resulting congestion effect are almost irrelevant for aggregate fluctuations; it is the exogenous oil trade shocks that generate large fluctuations.

There are two interesting avenues for future, related research. The first is to embed our setup within a general equilibrium framework that accounts for the oil import-export linkages so as to explore the role of oil transportation in magnifying or dampening the economic consequences of oil market imperfections for oil trade and prices. This initiative is inspired by Bornstein et al. (2022) where they build a structural model of the oil industry with the Organization of the Petroleum Exporting Countries (OPEC) acting as a cartel while other non-OPEC firms as a competitive fringe. In their model, despite its usefulness in helping us understand the strategic interactions between oil producing firms, it remains silent on any adjusting mechanism arising from the interactions between oil tankers and exporters as they only assume the traditional oil market clearing condition.

The second is closely related to the recent supply chain disruption triggered by the COVID-19 pandemic. The relentless surge in shipping prices and persistent bottlenecks at ports around the world have added to the barrage of problems affecting supply chains and sparked a growing discussion on their macroeconomic consequences.[4] Our empirical methodology and the resulting shipping dataset could be easily extended to visualize such issues, and the theoretical framework could be generalized so as to accommodate more industry-specific features (e.g., geographic dispersion of trade activities) and problem-specific setups (e.g., hysteresis, prolonged economic recovery, or stagflation).

[4]For instance, see https://www.ft.com/content/2daa10e0-4bf1-47e1-b62d-32adf5d5b44c (Accessed: 13 September 2021) for an observation related to the increasing inflation pressure on emerging economies, and https://www.ft.com/content/6c127633-9825-4921-9a11-71518591e2f3 (Accessed: 13 September 2021) for a trend towards regionalizing supply chains for multinational retailers and manufacturers.

Appendices

A

Appendices to "The Agricultural Exodus in the Philippines: Are Wage Differentials Driving the Process?"

A.1 Industry Classification

Since 2012, the Philippine Standard Industrial Classification (PSIC) has adopted the 2009 standard which changed significantly from the 1994 PSIC. Using the crosswalks, we create 16 major industry codes that are consistent over time. The classification is relatively coarse, but creating consistent codes necessarily requires aggregation of underlying two-digit codes into fairly broad categories.

The industries are: agriculture, forestry, and fishing; mining and quarrying; manufacturing; electricity, gas, and water supply; construction; wholesale and retail trade; repair of motor vehicles, motorcycles, and personal and household goods; hotels and restaurants; transport, storage and communications; financial intermediation; real estate, renting, and business activities; public administration, defense, and compulsory social security; education; health and social work; other community, social, and personal service activities; activities of private households as employers and undifferentiated production activities of private households; and extra-territorial organizations and bodies.

A.2 Theoretical Model

The goal of this section is to build a tractable model of the labor market capturing the conditions workers face when choosing employment status and industry. For simplicity, the model includes only two labor force statuses: employment (e) and non-employment (n); and two types of industries: agriculture (A) and non-agriculture (NA). To capture the empirical observation that job openings and job seekers simultaneously exist, we build a two-market random search model where labor market frictions prevent openings and job seekers from perfectly matching up. The model is motivated by the work of Den Haan et al. (2005) and Fujita (2018). We use the model to study the job separation and finding behaviors by agricultural workers, and consider the potential mechanisms contributing to the shrinking employment in agriculture. In what follows, we suppress the time indices until they become necessary for clarity.

A.2.1 Environment

The economy is populated by a unit mass of risk-neutral workers and an infinite mass of identical and infinitely lived firms. Each worker is endowed with one unit of labor, while each firm can hire at most one worker each time. When the job position is filled, the match produces output according to the following production function:

$$y_j(x_j) = x_j, j \in \{A, NA\} \tag{A.1}$$

where x_j is the industry-specific idiosyncratic productivity. When the match is first formed, it draws its idiosyncratic productivity from $G_j(x_j)$ with support $(0, \infty)$.

There are two separate matching markets for each type of industry and workers participate in the matching market for their type while looking for a job. For industry j, the matching process between workers and firms is formally depicted by the existence of a constant returns to scale matching function:

$$m_j(n_j, v_j) = \bar{m}_j n_j^{\alpha_j} v_j^{1-\alpha_j}, \tag{A.2}$$

where $\bar{m}_j$ stands for matching efficiency, α_j for the elasticity of the matching function with respect to non-employment, n_j for the measure of the non-employed, and v_j for the measure of vacancies. Let $\theta_j \equiv \frac{v_j}{n_j}$ denote labor market tightness. The job meeting rate can be written as $f_j(\theta_j) \equiv \frac{m_j}{n_j} = \bar{m}_j \theta_j^{1-\alpha_j}$, whereas the vacancy meeting rate is defined by $q_j(\theta_j) \equiv \frac{m_j}{v_j} = \bar{m}_j \theta_j^{-\alpha_j}$. The job and vacancy meeting rates satisfy the standard relationship $f_j(\theta_j) = q_j(\theta_j)\theta_j$.

We assume that workers face several possibilities in each period: firstly, both types of employed workers face the possibility that their productivities switch to a new level. The switch occurs with probability λ. When it occurs, the new productivity is drawn from $G_j(.)$. Each match can be endogenously terminated when the new productivity is too low. Secondly, workers switch industries with an exogenous probability, $1 - \gamma_j$.[1] Employed workers take into account the switching probability when deciding whether or not to endogenously separate, whilst non-employed workers consider it when deciding whether or not to remain in the same industry.

A.2.2 Value Functions

A.2.2.1 Workers

For industry $j \in \{A, NA\}$, let $W_j(x_j)$ denote the value of the j-type employed worker, written as:

$$\begin{aligned} W_j(x_j) = w_j(x_j) + \beta\Big(& (1-\lambda) W_j(x_j) \\ & + \lambda \mathbb{E}_{x'_j} \Big\{ \max \Big[W_j(x'_j), \gamma_j N_j + (1-\gamma_j) N_{\neg j} \Big] \Big\} \Big), \end{aligned} \tag{A.3}$$

where $w_j(x_j)$ is the current-period wage, β is the discount factor, x'_j is the productivity of this match in the next period, N_j is the value of the j-type non-employed worker, and $N_{\neg j}$ is the value of the non-employed worker in the other industry. The present value of being an employed worker in industry j is the earned

[1] To model the worker's industry switching behavior, we could apply an endogenous mechanism, e.g., assuming that workers draw some idiosyncratic industry-specific taste shocks which impact their switching decision (e.g., Artuç et al. (2010), Kennan and Walker (2011), and Pilossoph (2014)). However, we opt for the exogenous setting because it significantly reduces non-necessary analytical complexities without sacrificing any insight when compared to the endogenous counterpart.

wage w_j plus the continuation value. With probability $1-\lambda$, the worker continues its employment as productivity stays the same. With probability λ, productivity switches, and the worker decides whether or not to endogenously separate. The max operator in Equation (A.3) characterizes the optimal continuation/separation decision. Conditional on separation, with probability γ_j, the worker becomes non-employed in the current industry; with probability $1-\gamma_j$, the worker switches to the other industry and searches for jobs there. Note that we do not model job-to-job transitions in the current setting as in the LFS, job-to-job transitions only represent a small proportion of the flows between e_A and e_{NA} (2.4% and 1% for agricultural and non-agricultural workers respectively).

The value of the j-type non-employed worker is given by:

$$\begin{aligned} N_j = b_j + \beta\Big(f_j(\theta_j)\mathbb{E}_{x'_j}\Big\{ \max\Big[W_j(x'_j), \gamma_j N_j + (1-\gamma_j)N_{\neg j}\Big]\Big\} \\ + (1-f_j(\theta_j))\Big[\gamma_j N_j + (1-\gamma_j)N_{\neg j}\Big]\Big), \end{aligned} \tag{A.4}$$

where b_j is the flow outside value for the worker and $f_j(\theta_j)$ is the job meeting rate in industry j. A non-employed j-type worker earns a current period return of b_j, plus the expected continuation value. With probability $f_j(\theta_j)$, the worker meets a job and then decides whether to separate. Again, the max operator characterizes the optimal continuation/separation decision. With probability $1-f_j(\theta_j)$, the worker does not meet a job and continues to remain non-employed in the current industry with probability γ_j.

A.2.2.2 Jobs

For industry $j \in \{A, NA\}$, the job position filled with an employed worker embodies the following value:

$$J_j(x_j) = x_j - w_j(x_j) + \beta\Big\{(1-\lambda)J_j(x_j) + \lambda\mathbb{E}_{x'_j}\Big[\max(J_j(x'_j), V_j)\Big]\Big\}, \tag{A.5}$$

where V_j is the value of the unfilled position and the max operator characterizes the match continuation/destruction decision. The firm earns x_j, but must pay the worker a wage w_j. With probability $1-\lambda$, the employment continues as productivity

stays the same. With probability λ, productivity switches. It may be too low and thus the match may be rejected. The firm then starts the next period with an unfilled position, which has value equal to:

$$V_j = -c_j + \beta\Big\{q_j(\theta_j)\mathbb{E}_{x'_j}\Big[\max(J_j(x'_j), V_j)\Big] + (1 - q_j(\theta_j)V_j\Big\},$$

where c_j is the per period vacancy posting cost and $q_j(\theta_j)$ is the vacancy meeting rate. For simplicity, we assume free entry into the matching market that drives the value of the unfilled position to zero and thus the following job creation condition holds:

$$\frac{c_j}{\beta q_j(\theta_j)} = \mathbb{E}_{x'_j}\Big[\max(J_j(x'_j), 0)\Big]. \tag{A.6}$$

A.2.3 Surplus Sharing and Separation Decision

When the two parties meet, they either enter into an agreement, or they both revert to their outside options. The outside option of the worker is to either remain unmatched in the current industry or to switch to the other industry. The outside option of the firm is to remain unmatched and wait for another worker. The surplus of the match over the parties' outside options is split via generalized Nash bargaining.

When the employment relationship continues, each match in industry $j \in \{A, NA\}$ enjoys the surplus:

$$S_j(x_j) = J_j(x_j) + W_j(x_j) - \Big[\gamma_j N_j + (1 - \gamma_j)N_{\neg j}\Big].$$

The worker takes a constant fraction, denoted as $\eta \in (0, 1)$, of the total surplus and the firm takes the rest $1 - \eta$, i.e., $\eta J_j(x_j) = (1 - \eta)\Big\{W_j(x_j) - \Big[\gamma_j N_j + (1 - \gamma_j)N_{\neg j}\Big]\Big\}$. Given this rule, the worker and the firm agree on the continuation/separation decision. Since $J_j(x_j) + W_j(x_j)$ is increasing in x_j, there exists a cutoff productivity $\underline{x}_j$ below (above) which both sides choose to sever (continue) the employment relationship; at $\underline{x}_j$,

$$S_j(\underline{x}_j) = 0. \tag{A.7}$$

A.2.4 Partial Equilibrium in the Agricultural Labor Market

Having defined the value functions and separation decisions, we turn to characterize the partial equilibrium in the agricultural labor market. We derive a positively-sloped job separation curve and a negatively-sloped job creation curve in a plane comprised of reservation productivity $\underline{x}_A$ and labor market tightness θ_A. We use the framework to illustrate how in theory the 'push' and 'pull' factors in general affect the job separation and finding rates in the agricultural industry.

A.2.4.1 Job Separation

For simplicity, we define the expected surplus as $\mathbb{E}S_j(x'_j) \equiv \int_{\underline{x}_j}^{\infty} S_j(x'_j)dG_j(x'_j)$ for $j \in \{A, NA\}$. Combining Equations (A.3), (A.4), and (A.5), the evolution of the surplus for the A-type match is given by:

$$\begin{aligned} S_A(x_A) =& x_A - \Big[\gamma_A b_A + (1-\gamma_A) b_{NA}\Big] + \beta\Big[(1-\lambda) S_A(x_A) + \lambda \mathbb{E}S_A(x'_A)\Big] \\ & - \beta\eta\Big[\gamma_A f_A(\theta_A)\mathbb{E}S_A(x'_A) + (1-\gamma_A) f_{NA}(\theta_{NA})\mathbb{E}S_{NA}(x'_{NA})\Big] \\ & - \beta(1-\gamma_A)(1-\gamma_A-\gamma_{NA})(N_A - N_{NA}), \end{aligned} \tag{A.8}$$

where the difference in non-employment values, $N_A - N_{NA}$, is given by:

$$N_A - N_{NA} = \frac{b_A - b_{NA} + \beta\eta\Big(f_A(\theta_A)\mathbb{E}S_A(x'_A) - f_{NA}(\theta_{NA})\mathbb{E}S_{NA}(x'_{NA})\Big)}{1+\beta(1-\gamma_A-\gamma_{NA})}.$$

Evaluating Equation (A.8) at $\underline{x}_A$ yields the job separation condition in the agricultural labor market:

$$\begin{aligned} 0 =& \gamma_A b_A + (1-\gamma_A) b_{NA} + \frac{\beta(1-\gamma_A)(1-\gamma_A-\gamma_{NA})}{1+\beta(1-\gamma_A-\gamma_{NA})}(b_A - b_{NA}) \\ & - \beta\lambda\mathbb{E}S_A(x'_A) + \beta\eta\Big[\gamma_A f_A(\theta_A)\mathbb{E}S_A(x'_A) + (1-\gamma_A) f_{NA}(\theta_{NA})\mathbb{E}S_{NA}(x'_{NA})\Big] \\ & + \beta(1-\gamma_A)(1-\gamma_A-\gamma_{NA})\frac{\beta\eta\Big(f_A(\theta_A)\mathbb{E}S_A(x'_A) - f_{NA}(\theta_{NA})\mathbb{E}S_{NA}(x'_{NA})\Big)}{1+\beta(1-\gamma_A-\gamma_{NA})} - \underline{x}_A. \end{aligned} \tag{A.9}$$

Next, we add Equation (A.9) to Equation (A.8) and derive the surplus for the A-type match:

$$S_A(x_A) = \frac{x_A - \underline{x}_A}{1-\beta(1-\lambda)}.$$

By further assuming idiosyncratic productivity x_A is log-normally distributed with mean $\bar{x}_A$ and standard deviation σ, we can write the integral associated with truncated log-normal distribution in $\mathbb{E}S_A(x'_A)$ as:

$$\begin{aligned}\mathbb{E}S_A(x'_A) &\equiv \int_{\underline{x}_A}^{\infty} \frac{x'_A - \underline{x}_A}{1-\beta(1-\lambda)} dG_A(x'_A) \\ &= \frac{1}{1-\beta(1-\lambda)} \int_{\underline{x}_A}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\log(x'_A)-\bar{x}_A)^2} dx'_A - \frac{\underline{x}_A}{1-\beta(1-\lambda)}(1-G_A(\underline{x}_A)).\end{aligned}$$

Define $z \equiv \frac{\log(x'_A)-(\bar{x}_A+\sigma^2)}{\sigma}$. Using the substitution, $x'_A = e^{\bar{x}_A+\sigma^2}e^{\sigma z}$, and by change of variables, we derive the expected surplus for A-type match as:

$$\begin{aligned}\mathbb{E}S_A(x'_A) =& \frac{1}{1-\beta(1-\lambda)} e^{\bar{x}_A+\frac{1}{2}\sigma^2}\Big\{1-\Phi\Big[\frac{\log(\underline{x}_A)-(\bar{x}_A+\sigma^2)}{\sigma}\Big]\Big\} \\ &- \frac{\underline{x}_A}{1-\beta(1-\lambda)}(1-G_A(\underline{x}_A)),\end{aligned} \tag{A.10}$$

where $\Phi(.)$ and $G_A(.)$ represent the standard normal and log-normal cumulative distribution functions respectively. As such, we apply the implicit function theorem and summarize the relationship between reservation productivity $\underline{x}_A$ and tightness measure θ_A that is embedded in the job separation condition in the following proposition:

Proposition 1. *Define the job separation condition $F : \mathbb{R}^2 \to \mathbb{R}$ as in Equation* (A.9)*. Assuming the existence of a partial equilibrium such that a combination $(\underline{x}_{A,0}, \theta_{A,0}) \in \mathbb{R}^2$ yields $F(\underline{x}_{A,0}, \theta_{A,0}) = 0$ and some calibration such that $\frac{\partial F}{\partial \underline{x}_A}(\underline{x}_{A,0}, \theta_{A,0}) \neq 0$,*

$$\frac{d\underline{x}_A}{d\theta_A}(\underline{x}_{A,0}, \theta_{A,0}) > 0.$$

Proof. Given the stated assumptions, by the implicit function theorem, there is a neighborhood of $(\underline{x}_{A,0}, \theta_{A,0})$ such that whenever θ_A is sufficiently close to $\theta_{A,0}$, there is a unique $\underline{x}_A$ so that $F(\underline{x}_A, \theta_A) = 0$. This assignment makes $\underline{x}_A$ a continuous function of θ_A. As such, applying implicit differentiation to Equation (A.9) around $(\underline{x}_{A,0}, \theta_{A,0})$ yields:

$$\frac{d\underline{x}_A}{d\theta_A} = -\frac{\partial F/\partial \theta_A}{\partial F/\partial \underline{x}_A}. \tag{A.11}$$

In terms of $\frac{\partial F}{\partial \theta_A}$, we derive that:

$$\frac{\partial F}{\partial \theta_A} = \frac{\gamma_A(1-\beta)+\beta(1-\gamma_{NA})}{1+\beta(1-\gamma_A-\gamma_{NA})}\beta\eta\mathbb{E}S_A(x'_A)(1-\alpha_A)\bar{m}_A\theta_A^{-\alpha_A} > 0, \tag{A.12}$$

for $\theta_A \in (0,\infty)$. On the other hand, $\frac{\partial F}{\partial \underline{x}_A}$ is written as:

$$\frac{\partial F}{\partial \underline{x}_A} = \big[-\beta\lambda + \beta\eta f_A(\theta_A)\frac{\gamma_A+\beta(1-\gamma_A-\gamma_{NA})}{1+\beta(1-\gamma_A-\gamma_{NA})}\big]\frac{\partial \mathbb{E}S_A(x'_A)}{\partial \underline{x}_A} - 1.$$

Given Equation (A.10), we derive that $\frac{\partial \mathbb{E}S_A(x'_A)}{\partial \underline{x}_A} = -\frac{1-G_A(\underline{x}_A)}{1-\beta(1-\lambda)}$. Replacing $\frac{\partial \mathbb{E}S_A(x'_A)}{\partial \underline{x}_A}$ with this expression yields:

$$\frac{\partial F}{\partial \underline{x}_A} = \big[\beta\lambda - \beta\eta f_A(\theta_A)\frac{\gamma_A+\beta(1-\gamma_A-\gamma_{NA})}{1+\beta(1-\gamma_A-\gamma_{NA})}\big]\frac{1-G_A(\underline{x}_A)}{1-\beta(1-\lambda)} - 1.$$

Clearly the partial derivative $\frac{\partial F}{\partial \underline{x}_A}$ depends on the values of $(\underline{x}_A, \theta_A)$. However, taking into account that $\frac{\partial f_A(\theta_A)}{\partial \theta_A} > 0$ and $\frac{\partial G_A(\underline{x}_A)}{\partial \underline{x}_A} > 0$, we summarize its value by considering its two limits:

1. $\theta_A \to 0$: it implies $f_A(\theta_A) \to 0$, and the partial derivative has an upper limit as:
$$\frac{\partial F}{\partial \underline{x}_A} \to \frac{-\beta\lambda G_A(\underline{x}_A)-(1-\beta)}{1-\beta(1-\lambda)} < 0,$$
for any $\underline{x}_A \in (0,\infty)$ such that $G_A(\underline{x}_A) \in (0,1)$;

2. $\theta_A \to \infty$: it implies $f_A(\theta_A) \to \infty$, and the partial derivative has a lower limit as:
$$\frac{\partial F}{\partial \underline{x}_A} \to -\infty < 0,$$
for any $\underline{x}_A \in (0,\infty)$ such that $G_A(\underline{x}_A) \in (0,1)$.

Therefore, we conclude that:

$$\frac{\partial F}{\partial \underline{x}_A} < 0, \tag{A.13}$$

and combining it with Equations (A.11) and (A.12) gives:

$$\frac{d\underline{x}_A}{d\theta_A} > 0.$$

□

A.2.4.2 Job Creation

As given by Equation (A.6), the job creation condition illustrates that in equilibrium, the marginal benefit of posting a vacancy for a firm, that is its share of the expected surplus for a match, must equal the marginal cost which directly links to the vacancy posting cost and the easiness of meeting a worker. The job creation condition in the agricultural labor market can therefore be re-written as:

$$0 = (1-\eta)\mathbb{E}S_A(x'_A) - \frac{c_A}{\beta q_A(\theta_A)}, \tag{A.14}$$

from which we summarize the relationship between reservation productivity $\underline{x}_A$ and labor market tightness θ_A in the following proposition:

Proposition 2. *Define the job creation condition* $H : \mathbb{R}^2 \to \mathbb{R}$ *as in Equation* (A.14). *Assuming the existence of a partial equilibrium such that a combination* $(\underline{x}_{A,0}, \theta_{A,0}) \in \mathbb{R}^2$ *yields* $H(\underline{x}_{A,0}, \theta_{A,0}) = 0$ *and some calibration such that* $\frac{\partial H}{\partial \underline{x}_A}(\underline{x}_{A,0}, \theta_{A,0}) \neq 0$,

$$\frac{d\underline{x}_A}{d\theta_A}(\underline{x}_{A,0}, \theta_{A,0}) < 0.$$

Proof. Similar to Proposition 1, given the stated assumptions, $\underline{x}_A$ is a continuous function of θ_A, and applying implicit differentiation yields $\frac{d\underline{x}_A}{d\theta_A} = -\frac{\partial H/\partial \theta_A}{\partial H/\partial \underline{x}_A}$. The partial derivative $\frac{\partial H}{\partial \theta_A}$ is calculated as:

$$\frac{\partial H}{\partial \theta_A} = -\frac{c_A}{\beta \bar{m}_A}\alpha_A \theta_A^{\alpha_A - 1} < 0,$$

for $\theta_A \in (0, \infty)$, whereas the partial derivative $\frac{\partial H}{\partial \underline{x}_A}$ is written as:

$$\frac{\partial H}{\partial \underline{x}_A} = -\frac{(1-\eta)(1-G_A(\underline{x}_A))}{1-\beta(1-\lambda)} < 0,$$

for $\underline{x}_A \in (0, \infty)$. Consequently, we derive that $\frac{d\underline{x}_A}{d\theta_A} < 0$. □

As shown in Figure 2.4, there exist a positively-sloped job separation (JS) curve and a negatively-sloped job creation (JC) curve in the $(\underline{x}_A, \theta_A)$ plane. The intersection, $(\underline{x}_{A,0}, \theta_{A,0})$, marks the partial equilibrium in the agricultural labor market, and any shift in the job separation or creation curve due to changes in labor

market conditions in either the agricultural or non-agricultural labor market will result in corresponding changes in $(\underline{x}_{A,0}, \theta_{A,0})$, and thus affect the job separation and finding rates for agricultural workers.

A.2.5 Comparative Statics

In this section we will study the interactions between the JS and JC curves in Figure 2.4 so as to examine the influence of relevant factors on the job separation and finding in the agricultural industry, and how they contribute to or hinder the shrinking agricultural employment. In particular, we focus on two sets of factors - 'push' factors which capture incentives for workers to leave agriculture, and 'pull' factors which capture incentives for workers to remain in agriculture. Through comparative statics, we further illustrate that different 'push' ('pull') factors may stimulate (impede) the agricultural exodus through distinct channels. Lastly, this section complements our findings in Section 2.2 by including more analytical details.

A.2.5.1 'Push' Factors

First of all, let's consider an exogenous increase in mean productivity of the non-agricultural industry, $\bar{x}_{NA}$, such that the outside option for workers switching from agriculture to non-agriculture becomes more favorable, e.g., $f_{NA}(\theta_{NA})\mathbb{E}S_{NA}(x'_{NA})$ rises. In order for the job separation condition (A.9) to hold, at given levels of θ_A, the reservation productivity $\underline{x}_A$ will rise.[2] This implies a leftward shift of the JS curve to JS' as depicted in Figure 2.5. Assuming all else remain constant, at the new equilibrium where the JS' and JC curves intersect, the reservation productivity increases to $\underline{x}_{A,1}$ whilst the agricultural labor market tightness drops to $\theta_{A,1}$. Since the separation rate in agriculture can be written as $\lambda G_A(\underline{x}_A)$ which is monotonically increasing in $\underline{x}_A$, higher $\underline{x}_A$ implies greater job separation in agriculture in equilibrium: since the outside option value for agricultural workers

[2]Since the coefficient of the term $f_{NA}(\theta_{NA})\mathbb{E}S_{NA}(x'_{NA})$, $\frac{\beta\eta(1-\gamma_A)}{1+\beta(1-\gamma_A-\gamma_{NA})}$, is positive in Equation (A.9), having known that $\frac{\partial F}{\partial \underline{x}_A} < 0$, we derive that $\frac{d\underline{x}_A}{d\left(f_{NA}(\theta_{NA})\mathbb{E}S_{NA}(x'_{NA})\right)} = -\frac{\partial F/\partial\left(f_{NA}(\theta_{NA})\mathbb{E}S_{NA}(x'_{NA})\right)}{\partial F/\partial \underline{x}_A} > 0$ while holding θ_A fixed.

increases as a result of the increase in $\bar{x}_{NA}$, the surplus to be shared between agricultural workers and firms will decline, giving rise to that only the higher realizations of x_A will be accepted. In terms of the job finding rate, it is given by $f_A(\theta_A)(1 - G_A(\underline{x}_A))$, which is monotonically increasing in θ_A and decreasing in $\underline{x}_A$. Therefore, the confluence of higher $\underline{x}_A$ and lower θ_A in equilibrium implies a lower job finding rate in agriculture: not only do agricultural workers become more 'picky' about job offers and decline more often, but also they face more difficulties in meeting a potential employer. In reality, such a productivity difference often manifests itself through widening wage differentials (as documented in Figure 2.7), therefore incentivizing agricultural workers to move away from their land and search for non-agricultural jobs instead.

Similar dynamics can arise from an exogenous increase in the flow outside value for non-agricultural workers b_{NA}, or an exogenous increase in the matching efficiency in the non-agricultural industry $\bar{m}_{NA}$, as they both behave to raise the outside option value faced by agricultural workers. Lastly, it is noted the increasing productivity difference between agriculture and non-agriculture may also result from lower $\bar{x}_A$ instead of higher $\bar{x}_{NA}$, which in certain scenarios (see Figure A.1 for an illustration) will lead to both the JS and JC curves shifting to the left, and yield similar outcomes to those with only the JS curve shifting to the left.

In addition to the aforementioned 'push' factors, a boost to matching efficiency in the agricultural industry $\bar{m}_A$ may also lead to significant employment outflows from agriculture, even though the exact mechanism through which the outflows take place varies from the previous ones. Specifically, an exogenous increase in $\bar{m}_A$ triggers not only a leftward shift of the JS curve, but also a rightward shift of JC curve.[3,4] The intuition is that the boost to matching efficiency in the agricultural industry not only enhances the outside option for its workers (as they can find jobs more easily),

[3]It can be shown from Equation (A.9) that the multiplicative terms before $f_A(\theta_A)$ write as $\frac{\gamma_A(1-\beta)+\beta(1-\gamma_{NA})}{1+\beta(1-\gamma_A-\gamma_{NA})}\beta\eta\mathbb{E}S_A(x'_A) > 0$. Therefore, given that $f_A(\theta_A) \equiv \bar{m}_A\theta_A^{1-\alpha_A}$, it implies at each $\underline{x}_A$ where Equation (A.9) holds, an increase in $\bar{m}_A$ results in a corresponding decline in θ_A such that the JS curve shifts to the left.

[4]From Equation (A.14) we can clearly see that at each $\underline{x}_A$ where it holds, for an exogenous increase in $\bar{m}_A$, θ_A will rise accordingly, provided that $q_A(\theta_A) \equiv \bar{m}_A\theta_A^{-\alpha_A}$.

but also improves the ability of firms in filling their vacancies and encourages them to create more agricultural jobs. Graphically, in Figure A.2, assuming that the shift in the JC curve is relatively small (which often occurs when α_A is small at low levels of θ_A), at the new equilibrium where JS''' and JC''' curves intersect, the reservation productivity increases to $\underline{x}_{A,3}$ whilst the agricultural labor market tightness drops to $\theta_{A,3}$. This leads to similar outcomes as in the other cases, namely, greater job separation but lower job finding probabilities in the agricultural industry. Therefore, given such model predictions, it is expected that better-functioning labor market and increasing measures to boost matching efficiency (e.g., improved transport infrastructure) will serve to promote agricultural employment outflows.

A.2.5.2 'Pull' Factors

Now we turn to the 'pull' factors and analyze how they influence the job separation and finding rates in agriculture. We start with an exogenous decrease in the flow outside value for agricultural workers, b_A. Holding all else constant, in order for the job separation condition (A.9) to hold, for given values of θ_A, this time the reservation productivity $\underline{x}_A$ will fall. This implies a rightward shift of the JS curve to JS^* as depicted in Figure 2.6. As a result, at the new equilibrium where JS^* and JC curves intersect, the reservation productivity decreases to $\underline{x}_{A,4}$ whilst the agricultural labor market tightness increases to $\theta_{A,4}$; compared to the original equilibrium, the job separation rate falls whilst the job finding rate increases in the agricultural industry.[5] The intuition is simple: the decrease in the agricultural worker's flow outside value enlarges the surplus to be shared, which then leads to that lower realizations of idiosyncratic productivities – which were previously rejected – are now accepted; in other words, workers become less 'picky' and become less likely to separate. In practice, the decrease in the flow outside value for agricultural workers is often associated with lack of education or skill depreciation,

[5]Again, in theory, similar dynamics can arise from an exogenous decrease in the flow outside value for non-agricultural workers b_{NA} or an exogenous decrease in the matching efficiency in the non-agricultural industry $\bar{m}_{NA}$, as both of them behave to shrink the outside option value faced by agricultural workers. However, we rarely find empirical or anecdotal evidence supporting these two arguments in practice.

as they cannot satisfy the skill requirements for non-agricultural jobs and thus get stuck in agriculture. A similar situation can also occur when there is a significant agricultural cluster effect due to lock-ins.

Another 'pull' factor to note is the deteriorating bargaining power of agricultural workers in their wage negotiation, which may also arise in the case of obsolete skills or agricultural clusters. Yet, the mechanism through which the decrease in η hinders agricultural exodus differs from that illustrated in Figure 2.6; it instead involves rightward shifts in both the JS and JC curves.[6] The dynamics are shown in Figure A.3, and the new equilibrium is where the JS^{**} and JC^{**} curves intersect, provided that the shift in the JC curve is relatively small (which often occurs when α_A is large at high levels of θ_A). The new equilibrium features lower job separation but higher job finding probabilities in the agricultural industry because on the one hand, the lower bargaining power of workers shrinks their share of the matching surplus and discourages job separation, while on the other hand, the resulting greater bargaining power of firms (i.e., $1-\eta$) increases their share of the matching surplus and promotes job creation. Hence, given the model predictions, it is expected that inadequate education and significant agricultural clusters will serve to dampen agricultural employment outflows.

[6] First, the terms associated with η in Equation (A.9) can be collected as $\eta\Big\{\frac{\gamma_A(1-\beta)+\beta(1-\gamma_{NA})}{1+\beta(1-\gamma_A-\gamma_{NA})}\beta f_A(\theta_A)\mathbb{E}S_A(x'_A)+\frac{(1-\gamma_A)[(1-\beta)+\beta(\gamma_A+\gamma_{NA})]}{1+\beta(1-\gamma_A-\gamma_{NA})}\beta f_{NA}(\theta_{NA})\mathbb{E}S_{NA}(x'_{NA})\Big\}$. Therefore, at each $\underline{x}_A$ where Equation (A.9) holds, a reduction in η will lead to a corresponding increase in θ_A, holding all else constant. On the other hand, from Equation (A.14), it implies that at each $\underline{x}_A$ while the condition is satisfied, a reduction in η will lead to a corresponding increase in θ_A.

A.3 End Behaviors of the EN and NE Rates

To explore on the extent to which our baseline regression results are affected by the nonconforming end behaviors of EN and NE rates as documented in Figures 2.3a and 2.3b, we apply two exercises, in which we add a time dummy indicating 2005Q2-2008Q1 and 2017Q1-2019Q4 respectively in the LPM of each labor market transition rate, and interact it with the proposed 'push' and 'pull' factors.

The estimated AMEs of each factor during different sample periods are presented in Tables A.1 and A.2. In Table A.1, we find that most results are robust, except that the AME of real wage differentials on $NE_{Nonagri}$ turns negative and highly significant during 2005Q2-2008Q1, which is in principle consistent with the stagnation of real wage differentials during the same period, as shown in Figure 2.7. On the other hand, in Table A.2, we find that the estimated effects of real wage differentials, labor market efficiency, high school and college education, and palay and corn production on the EN rates are robust in 2005Q2-2016Q4; however, increasing road density reduces job separations for both agricultural and non-agricultural workers, reflecting that road density may serve as a proxy for some other factor in the regression. The estimated effects of real wage differentials, labor market efficiency, high school, and college education on the NE rate for agricultural workers are robust in the limited sample period; in contrast, only the estimated effects of road density and college education on the NE rate for non-agricultural workers are robust in 2005Q2-2016Q4. Given the evidence, we conclude that most of the inconsistencies in results are associated with the behavior of $NE_{Nonagri}$ towards the end of the sample period, which may result from the implementation of the New Master Sample Design of the LFS starting in 2016Q2.

Table A.1: AMEs in the LPM with time dummy representing 2005Q2-2008Q1

	EN_{Agri}	$EN_{Nonagri}$	NE_{Agri}	$NE_{Nonagri}$
Real Wage Differentials				
2005Q2-2008Q1	0.0168***	-0.00603	-0.0224***	-0.0240***
	(0.00580)	(0.00546)	(0.00225)	(0.00356)
2008Q2-2019Q4	0.0106***	-0.00682***	-0.00812***	0.0171***
	(0.00308)	(0.00241)	(0.00128)	(0.00245)
Labor Market Efficiency				
2005Q2-2008Q1	-0.0111	-0.0133**	0.00411	0.0193***
	(0.00763)	(0.00625)	(0.00345)	(0.00427)
2008Q2-2019Q4	0.0407***	-0.0138**	0.00356	0.0586***
	(0.00698)	(0.00582)	(0.00328)	(0.00397)
Road Density				
2005Q2-2008Q1	0.152***	0.00601	-0.108***	0.0281
	(0.0257)	(0.0231)	(0.0136)	(0.0218)
2008Q2-2019Q4	0.151***	0.0101	-0.103***	0.0240
	(0.0253)	(0.0231)	(0.0135)	(0.0215)
High School Education				
2005Q2-2008Q1	0.0218***	0.00248***	-0.00884***	0.00364***
	(0.000953)	(0.000911)	(0.000480)	(0.000639)
2008Q2-2019Q4	0.0201***	0.00281***	-0.0216***	-0.00379***
	(0.000545)	(0.000537)	(0.000406)	(0.000564)
College Education				
2005Q2-2008Q1	0.0332***	-0.0110***	-0.0167***	0.00271***
	(0.00179)	(0.000906)	(0.000470)	(0.000699)
2008Q2-2019Q4	0.0316***	-0.0120***	-0.0293***	0.00138**
	(0.00109)	(0.000555)	(0.000395)	(0.000608)
Palay & Corn Production				
2005Q2-2008Q1	-0.172***	0.0419	0.0212	-0.125***
	(0.0359)	(0.0372)	(0.0151)	(0.0255)
2008Q2-2019Q4	-0.105***	-0.00510	0.00736	-0.0788***
	(0.0205)	(0.0208)	(0.0104)	(0.0199)
Observations	1437827	2553367	2448570	2496782

Notes:

(1) All standard errors (in parentheses) are heteroskedasticity robust;

(2) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$;

(3) Demographic, region, and year & quarter dummies are included.

Table A.2: AMEs in the LPM with time dummy representing 2017Q1-2019Q4

	EN_{Agri}	$EN_{Nonagri}$	NE_{Agri}	$NE_{Nonagri}$
Real Wage Differentials				
2005Q2-2016Q4	0.0164***	-0.000583	-0.00456***	-0.00452***
	(0.00250)	(0.00224)	(0.000818)	(0.00160)
2017Q1-2019Q4	0.00609	-0.00710	0.00458	0.0320***
	(0.00916)	(0.00552)	(0.00363)	(0.00628)
Labor Market Efficiency				
2005Q2-2016Q4	0.117*	-0.160***	-0.210***	-0.272***
	(0.0663)	(0.0381)	(0.0208)	(0.0462)
2017Q1-2019Q4	0.162***	-0.119***	-0.192***	-0.191***
	(0.0628)	(0.0345)	(0.0193)	(0.0428)
Road Density				
2005Q2-2016Q4	-0.172***	-0.0900***	0.00491	0.0656***
	(0.0289)	(0.0262)	(0.0120)	(0.0174)
2017Q1-2019Q4	-3.583***	-1.252*	-0.0503	-1.552**
	(1.077)	(0.723)	(0.565)	(0.739)
High School Education				
2005Q2-2016Q4	0.0199***	0.00328***	-0.00761***	0.00337***
	(0.000477)	(0.000467)	(0.000255)	(0.000339)
2017Q1-2019Q4	0.0231***	0.00100	-0.0549***	-0.0202***
	(0.00156)	(0.00131)	(0.00114)	(0.00159)
College Education				
2005Q2-2016Q4	0.0341***	-0.0109***	-0.0140***	0.00185***
	(0.000968)	(0.000479)	(0.000251)	(0.000367)
2017Q1-2019Q4	0.0225***	-0.0151***	-0.0694***	0.00163
	(0.00278)	(0.00135)	(0.00114)	(0.00179)
Palay & Corn Production				
2005Q2-2016Q4	-0.133***	-0.00649	0.00918	-0.000699
	(0.0167)	(0.0163)	(0.00651)	(0.0114)
2017Q1-2019Q4	-0.0402	-0.0850	-0.0514	-0.339***
	(0.0636)	(0.0587)	(0.0395)	(0.0722)
OObservations	1437827	2553367	2448570	2496782

Notes:

(1) All standard errors (in parentheses) are heteroskedasticity robust;

(2) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$;

(3) Demographic, region, and year & quarter dummies are included.

A.4 Additional Tables and Figures

Table A.3: AMEs in the LPM with clustered standard errors at the regional level

	EN_{Agri}	$EN_{Nonagri}$	NE_{Agri}	$NE_{Nonagri}$
'Push' Factors				
Real Wage Differentials	0.00947	-0.00572	-0.00683	0.00631
	(0.00683)	(0.00355)	(0.00540)	(0.0191)
Labor Market Efficiency	0.00675	-0.0175***	-0.00174	0.0496***
	(0.00927)	(0.00519)	(0.00580)	(0.00967)
Road Density	0.0487	-0.0170	-0.0768	0.0552
	(0.0985)	(0.0654)	(0.0889)	(0.0968)
'Pull' Factors				
High School Education	0.0204***	0.00276**	-0.0190***	-0.00237**
	(0.00281)	(0.000956)	(0.00365)	(0.00109)
College Education	0.0319***	-0.0118***	-0.0267***	0.00173
	(0.00454)	(0.00117)	(0.00517)	(0.00159)
Palay & Corn Production	-0.112*	-0.0208	0.00976	-0.0340
	(0.0583)	(0.0243)	(0.0530)	(0.0689)
Observations	1437827	2553367	2448570	2496782

Notes:
(1) All standard errors (in parentheses) are clustered at the regional level;
(2) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$;
(3) Demographic, region, and year & quarter dummies are included.

Table A.4: AMEs of the proposed 'push' and 'pull' factors in a logit model

	EN_{Agri}	$EN_{Nonagri}$	NE_{Agri}	$NE_{Nonagri}$
'Push' Factors				
Real Wage Differentials	0.0103***	-0.00555***	-0.00300***	0.00180
	(0.00243)	(0.00210)	(0.000947)	(0.00179)
Labor Market Efficiency	-0.00129	-0.0638***	0.00363***	-0.0102
	(0.00413)	(0.0114)	(0.00136)	(0.00741)
Road Density	0.0393**	-0.00503	-0.0233***	0.0678***
	(0.0189)	(0.0182)	(0.00901)	(0.0164)
'Pull' Factors				
High School Education	0.0189***	0.00183***	-0.0173***	-0.00208***
	(0.000468)	(0.000510)	(0.000331)	(0.000452)
College Education	0.0295***	-0.0123***	-0.0258***	0.000924*
	(0.000946)	(0.000520)	(0.000318)	(0.000489)
Palay & Corn Production	-0.112***	-0.0145	-0.00166	-0.0136
	(0.0165)	(0.0170)	(0.00731)	(0.0153)
Observations	1437827	2553367	2448570	2496782

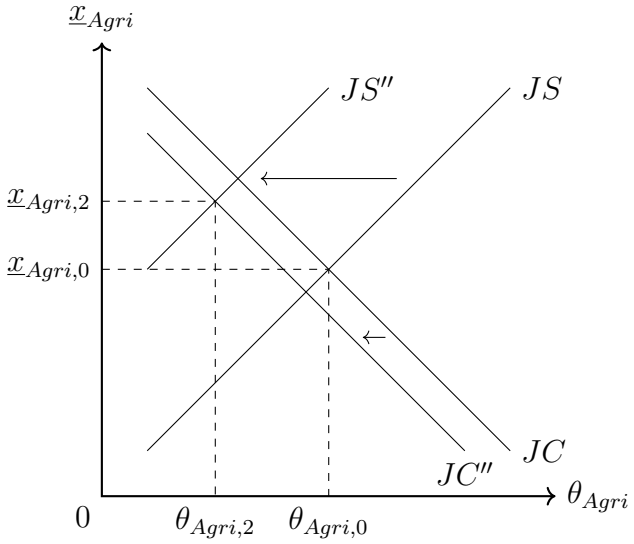
Notes:

(1) All standard errors (in parentheses) are clustered at the regional level;

(2) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$;

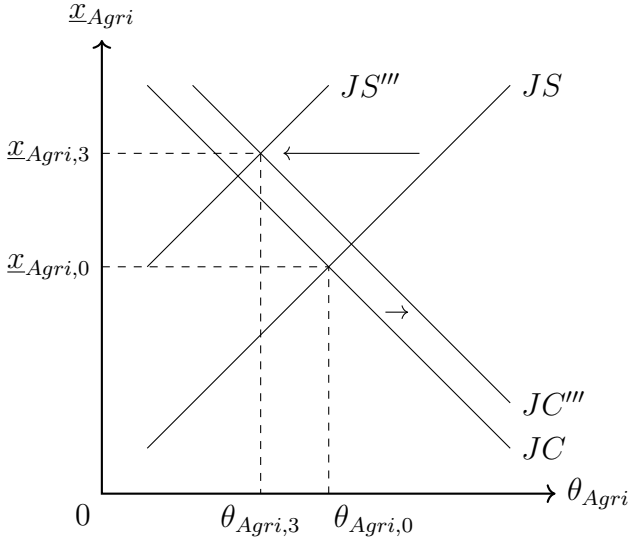
(3) Demographic, region, and year & quarter dummies are included.

Figure A.1: The influence of lower agricultural productivity on job separation and finding



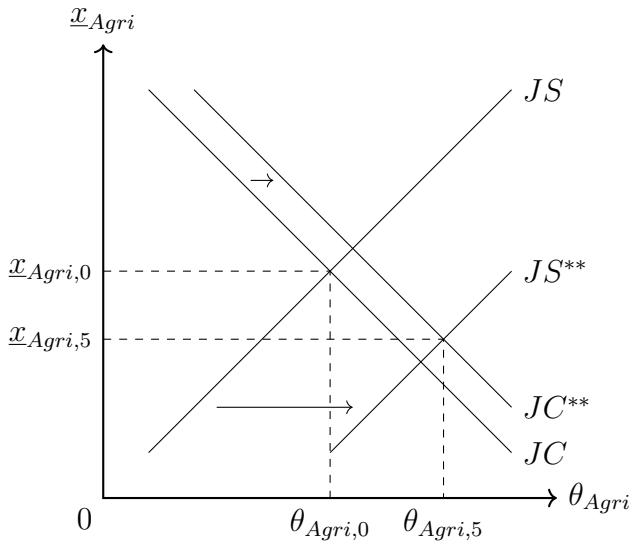
Notes: (1) JS'' and JC'' represent the shifted job separation and creation curves following an exogenous decrease in the agricultural productivity; (2) the intersection between the JS'' and JC'' curves represents the new (partial) equilibrium in the agricultural labor market, featuring $(\underline{x}_{Agri,2}, \theta_{Agri,2})$.

Figure A.2: The influence of better matching in agriculture on job separation and finding



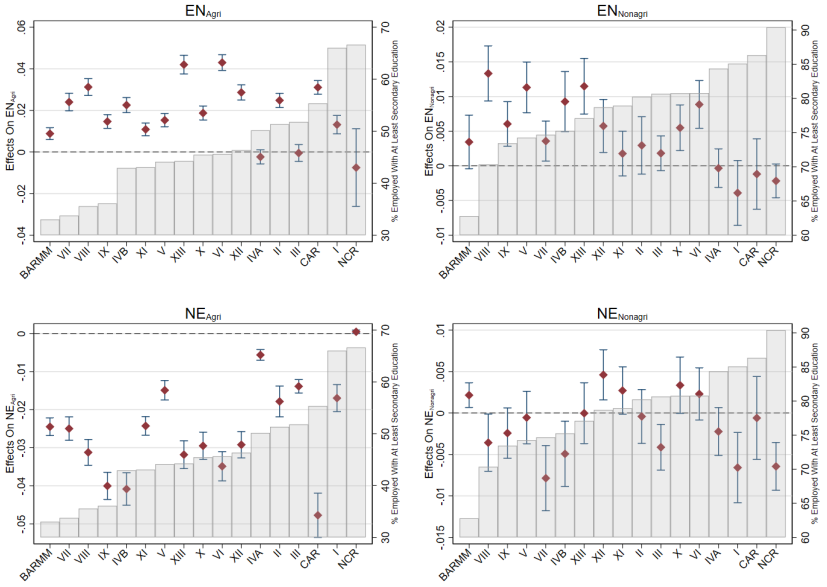
Notes: (1) JS''' and JC''' represent the shifted job separation and creation curves following an exogenous increase in the matching efficiency in the agricultural industry; (2) the intersection between the JS''' and JC''' curves represents the new (partial) equilibrium in the agricultural labor market, featuring $(\underline{x}_{Agri,3}, \theta_{Agri,3})$.

Figure A.3: The influence of weaker bargaining of agricultural workers on job separation and finding



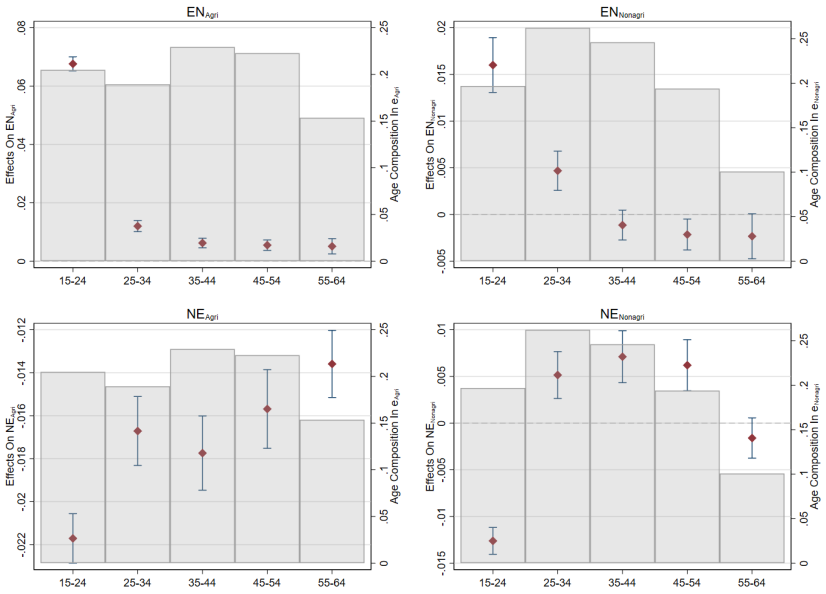
Notes: (1) JS^{**} and JC^{**} represent the shifted job separation and creation curves following an exogenous decrease in the bargaining power of agricultural workers; (2) the intersection between the JS^{**} and JC^{**} curves represents the new (partial) equilibrium in the agricultural labor market, featuring $(\underline{x}_{Agri,5}, \theta_{Agri,5})$.

Figure A.4: AMEs of high school education, by region



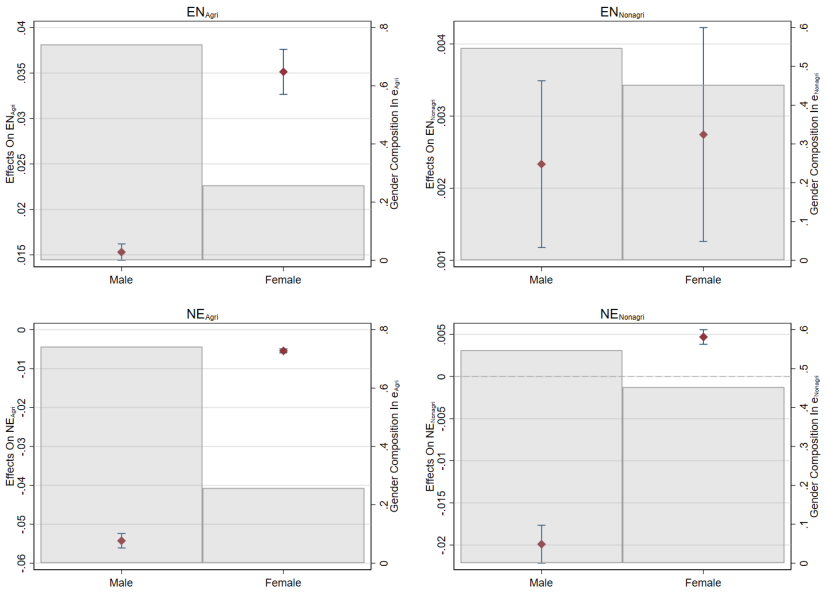
Source: Labor Force Survey, Philippine Statistics Authority.

Figure A.5: AMEs of high school education, by age



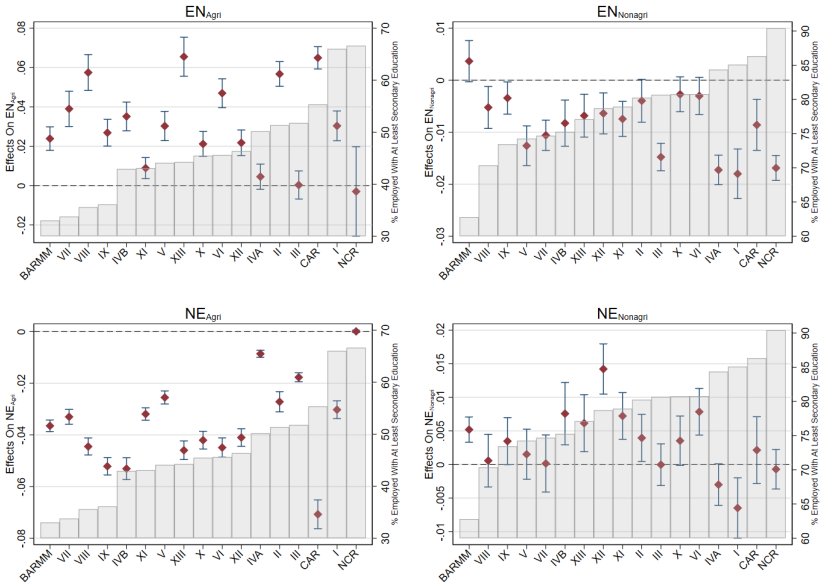
Source: Labor Force Survey, Philippine Statistics Authority. The underlying LPM includes interaction terms between high school education dummy, region dummies, and age dummies.

Figure A.6: AMEs of high school education, by gender



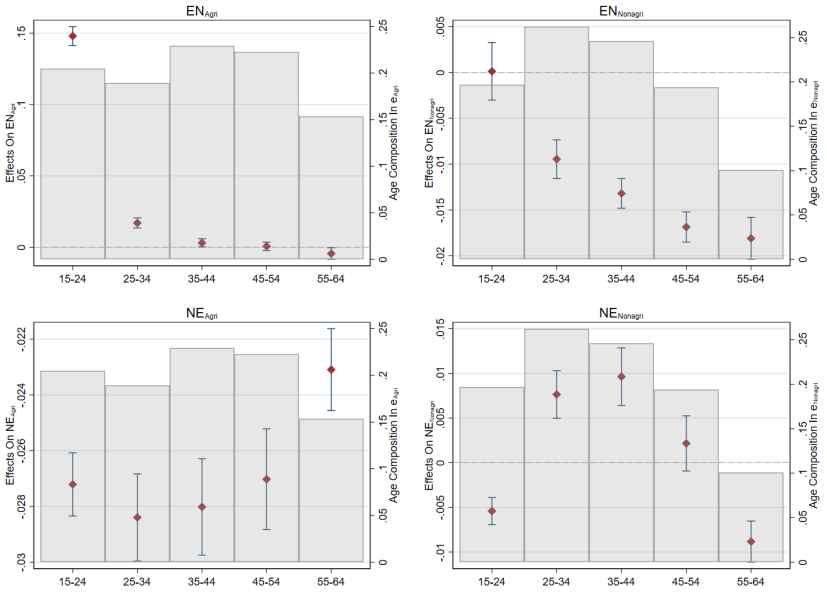
Source: Labor Force Survey, Philippine Statistics Authority. The underlying LPM includes interaction terms between high school education dummy, region dummies, and gender dummies.

Figure A.7: AMEs of college education, by region



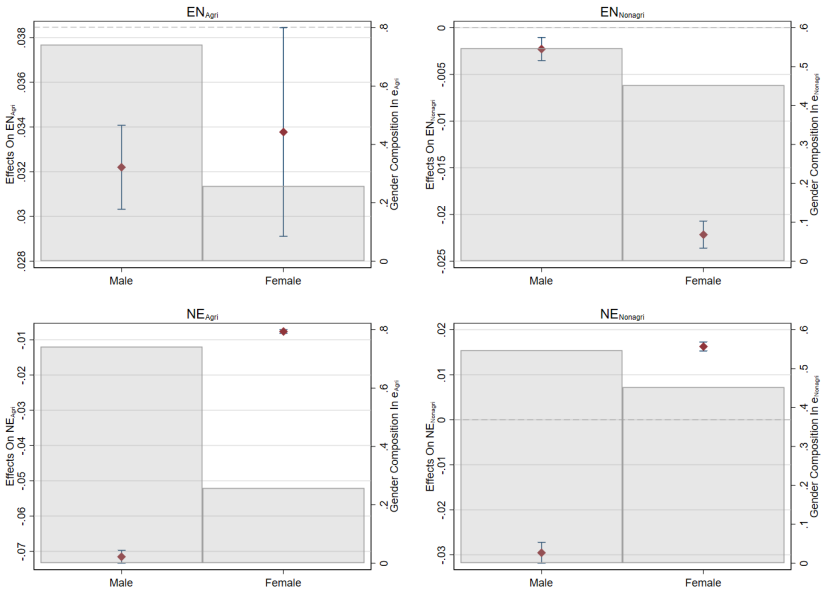
Source: Labor Force Survey, Philippine Statistics Authority.

Figure A.8: AMEs of college education, by age



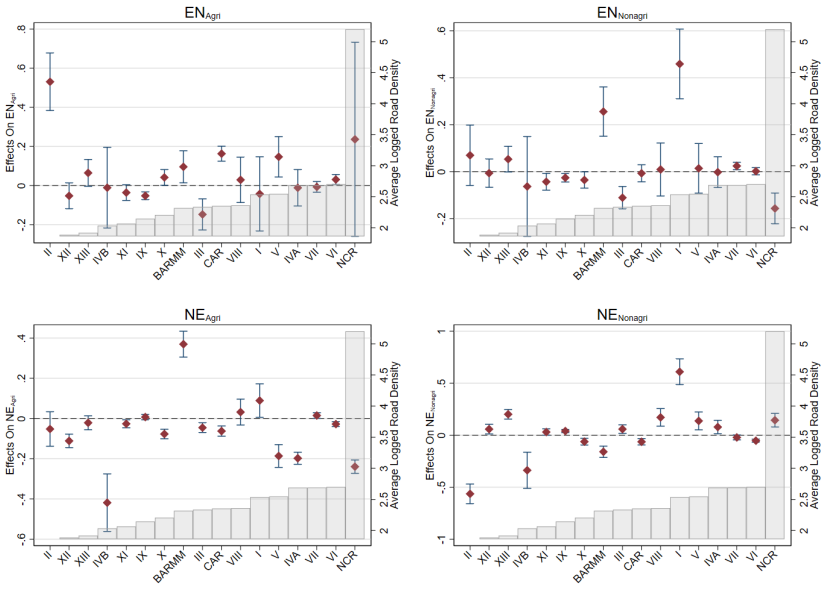
Source: Labor Force Survey, Philippine Statistics Authority. The underlying LPM includes interaction terms between college education dummy, region dummies, and age dummies.

Figure A.9: AMEs of college education, by gender



Source: Labor Force Survey, Philippine Statistics Authority. The underlying LPM includes interaction terms between college education dummy, region dummies, and gender dummies.

Figure A.10: AMEs of road density, by region



Source: Labor Force Survey, Philippine Statistics Authority; 2018 & 2019 ATLAS, Department of Public Works and Highways.

B

Appendices to "A Big Data Approach to Cargo Type Prediction and Its Implications for Oil Trade Estimation"

B.1 Additional Tables and Figures

Table B.1: Summary statistics of input variables (2nd stage)

Input variable	Unit	Mean	S.D.	Median	Min.	Max.
Geography						
Trip distance	Nautical mile	2523.3	2418.8	1646.9	60.2	12402.6
Proportion of crude oil trips from the origin country	Percent	64.5	21.0	71.4	0.0	100.0
Proportion of crude oil trips to the destination country	Percent	64.9	20.2	62.1	0.0	100.0
Shipowner/vessel/operation heterogeneity						
Proportion of crude oil trips by the shipowner	Percent	64.8	19.4	68.6	0.0	100.0
Vessel's type	N/A	0.41	0.49	0.00	0.00	1.00
Vessel's age	Year	13.2	4.4	13.0	4.0	33.0
In-port duration	Day	6.3	5.5	4.0	1.0	29.0

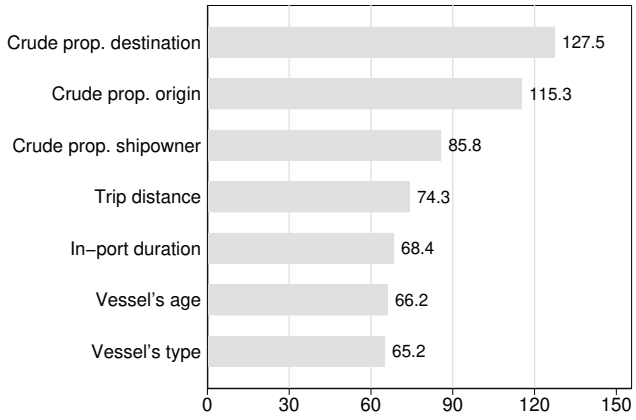
Notes. The underlying sample from which the summary statistics are calculated comprises the set of estimated DPP trips from the first stage. The proportion of crude oil trips from the country of origin (of an estimated DPP trip) is calculated as the proportion of crude oil trips originating from the estimated DPP trip's loading country, over the total number of DPP trips initiated from that country. The proportion of crude oil trips to the country of destination (of an estimated DPP trip) is calculated as the proportion of crude oil trips finishing at the estimated DPP trip's discharge country, over the total number of DPP trips ended at that country. The proportion of crude oil trips carried by the shipowner (of an estimated DPP trip) is calculated as the proportion of crude oil trips that the estimated DPP trip's designated shipowner engages in, over the total number of DPP trips completed by that shipowner. Each vessel's type is labeled as 0 if it belongs to the LR1 category, and 1 if it belongs to the LR2 category.

Table B.2: Summary of the prediction results on the test set (2nd stage)

Classifier	ACC	Sensitivity	Specificity	AUC
Random forest	76.5%	60.9%	83.4%	0.770
Logistic regression	73.8%	73.2%	74.1%	0.786
SVM	75.2%	72.5%	76.4%	0.785
Neural network	74.9%	60.9%	81.8%	0.790

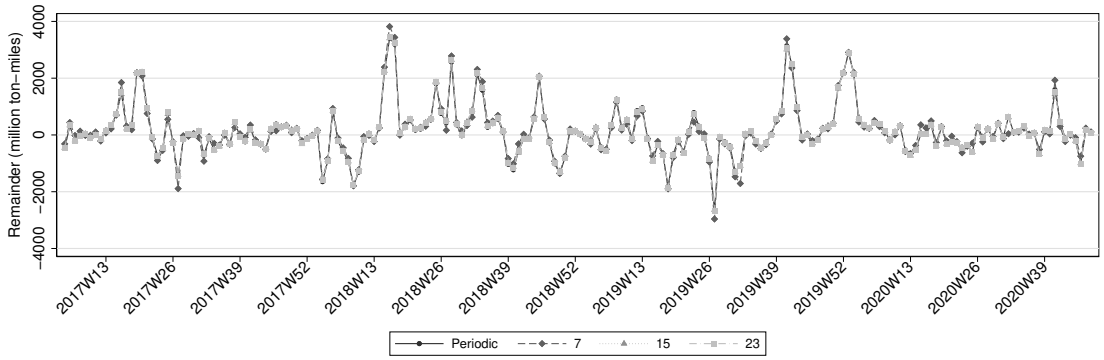
Notes. ACC measures a model's overall predictive accuracy, and is calculated as the proportion of cargo types that the model correctly classifies. Sensitivity measures the proportion of dirty refined product oil trips that are correctly classified, while specificity measures the correctly classified proportion of crude oil trips. AUC measures the ability of a model to distinguish between crude and dirty refined product oil trips.

Figure B.1: MDA values of the input variables (2nd stage)



Notes. Based on their MDA values, the selected input variables are ranked from greatest to least relative importance in classifying loaded trips into the correct classes.

Figure B.2: Remainder components of STL with different values of n_s



Notes. The figure plots the remainder component of each STL decomposition of the estimated global crude oil trade volume carried by LR tankers. The seasonal smoothing parameter n_s is set at different values. Our choices of n_s are based on the suggestion of Cleveland et al. (1990) that n_s should be odd, and no smaller than 7.

B.2 Alternative Prediction Models

In this section we lay out and briefly describe the underlying mechanism embedded in each of three alternative prediction models, namely, logistic regression, support vector machine (SVM), and neural network.

B.2.1 Logistic Regression

Logistic regression is one of the supervised learning models that is frequently employed in binary-class and multi-class classification problems, due to its simplicity and computational efficiency. In a binary classification problem, logistic regression can be written as follows:

$$\log\Big[\frac{P(Y=1|X=x)}{P(Y=0|X=x)}\Big] = w^T x + b,$$

where $P(Y=1|X=x)$ and $P(Y=0|X=x)$ are the conditional probabilities of the response variable Y in each class, and w and b are the coefficients of input variables X and an intercept, respectively.

We run the logistic regression with the CPP trips labeled as the positive class. Table B.3 summarizes the estimated coefficients of the input variables and their associated p-values. The signs of most coefficients align with our expectations, as

Table B.3: Parameter estimates of the logistic regression model

Coefficient	**Estimate**	p**-value**
Intercept	0.059	0.393
Geography		
Trip distance	0.834	0.000
Proportion of CPP trips from the origin country	1.341	0.000
Proportion of CPP trips to the destination country	1.179	0.000
Market		
Freight earnings spread	0.097	0.269
Short-term forward yield differences	-0.221	0.136
Medium-term forward yield differences	0.369	0.019
Long-term forward yield differences	-0.090	0.248
Shipowner/vessel/operation heterogeneity		
Proportion of CPP trips carried by the shipowner	1.697	0.000
Vessel's type	0.713	0.000
Vessel's age	-0.617	0.000
In-port duration	-0.084	0.219

Notes. We run the logistic regression with CPP trips labeled as the positive class. In addition, z-score normalization is applied to the data preprocessing.

discussed in the main text. For example, the positive sign of the coefficient of trip distance indicates that the probability that a ship carries CPP increases with its trip distance, which is consistent with the pattern in Figure 3.2. In addition, the positive sign of the coefficient of a vessel's age supports our proposition that younger tankers are more technically suitable for CPP transportation, and are preferred by charterers that trade CPP.

More importantly, the coefficients of trip distance and the proportion of CPP trips for the countries of origin and destination of a loaded trip are statistically significant at 1% level of significance. This further indicates that geography plays a critical role in determining the distribution of CPP and DPP trips around the globe. Likewise, the coefficient of the proportion of CPP trips carried by the shipowner of a loaded trip, along with those related to the technical specifications of the employed vessel, are also statistically significant at 1% level of significance. This points to the importance of shipowner and vessel heterogeneity in estimating the cargo types. Finally, the market-related input variables (except for the medium-term forward

yield differences) seem irrelevant to the determination of cargo type, which implies that the fluctuations of shipping markets do not necessarily impact the shipowners' choice of where to search for oil delivery jobs.

B.2.2 SVM

SVM is another popular classifier in the domain of supervised machine learning. It can be expressed as the following problem:

$$\min_{w,\epsilon} \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{N}\epsilon_i$$
$$s.t.,$$
$$\epsilon_i \geq 0, \forall i \in N,$$
$$y_i(x_i^T w + b) \geq 1 - \epsilon_i, \forall i \in N,$$

where w and b represent the coefficients of the decision boundary, C the penalty coefficient, and ϵ_i the violation of the hard margin constraint of each data point. In addition, y_i is either 1 or -1, each of which values indicates the class to which point x_i belongs. In our classification problem, when applying the SVM, we employ z-score normalization and grid search, the latter of which is adapted to tune the hyper-parameter C. We also use the Gaussian radial basis kernel to expand the dimensions of data and to help improve the performance of the SVM.

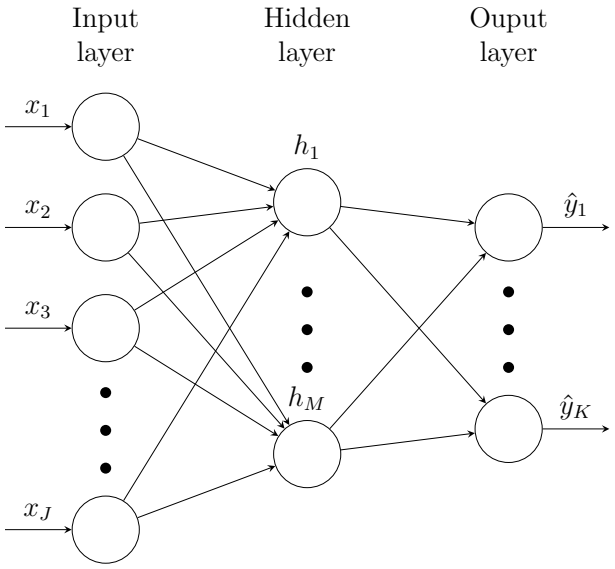
B.2.3 Neural Network

We also explore the application of a neural network in estimating the cargo types.[1] As shown in Figure B.3, a neural network takes a vector of input variables and transforms them into a vector of outputs, each representing the probability that an observation falls in a certain class.

What makes the neural network unique compared to other supervised learning models is the inclusion of a hidden layer. Specifically, a neuron in the hidden layer (h_m) takes the input vector $(x_1, \ldots, x_J)$, and transforms it to a scalar by

[1]This is underpinned by the Universality Approximation Theorem (Cybenko, 1989), which states that any continuous function can be approximated by a neural network with one hidden layer, a sufficient number of hidden neurons, and an appropriate activation function.

Figure B.3: Basic structure of a neural network



multiplying the vector by the corresponding weights and adding a bias term to the result. Then, the scalar is transformed to the output of a neuron in the hidden layer by an activation function. Applying the same procedures to each neuron in the output layer, the output probabilities ($\hat{y}$) is written as:

$$\hat{y} = \begin{pmatrix} \hat{y}_1 \\ \dots \\ \hat{y_K} \end{pmatrix} = \begin{pmatrix} f_y\Big\{ \sum_{m=1}^{M} \Big[w_{m1} f_h(\sum_{j=1}^{J} u_{jm} x_j + c_m) \Big] + b_1 \Big\} \\ \dots \\ f_y\Big\{ \sum_{m=1}^{M} \Big[w_{mK} f_h(\sum_{j=1}^{J} u_{jm} x_j + c_m) \Big] + b_K \Big\} \end{pmatrix},$$

where u_{jm} and c_m represent the weight attached to the j-th neuron in the input layer, and the bias term when calculating the output of the m-th neuron in the hidden layer; w_{mk} and b_k indicate the weight attached to the m-th neuron in the hidden layer, and the bias term when calculating the output of the k-th neuron in the output layer; J, M, and K denote the number of neurons in each of the input, hidden, and output layers; and, f_h, f_y give the activation functions of the hidden and output layers, respectively. Note that $K = 2$ in our binary classification problem.

Finally, in our application of the neural network, drop-out and early stopping are simultaneously adopted so as to avoid over-fitting. Hyper-parameters, including the number of nodes, the drop-out rate of the hidden layer, and an optimizer for the model, are tuned via grid search. Once again, z-score normalization is employed to preprocess the shipping data.

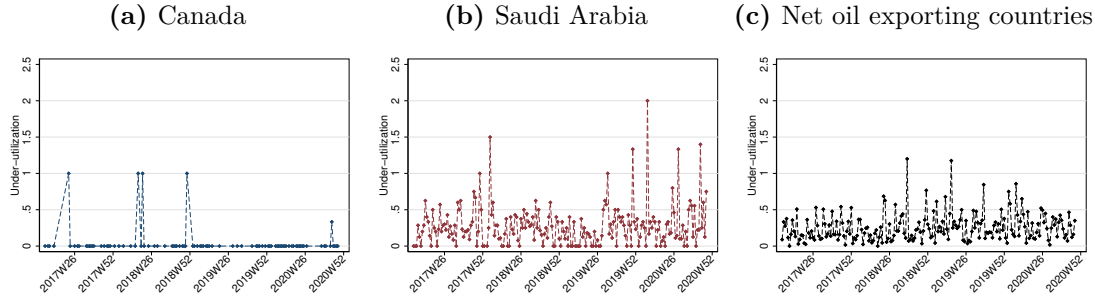
C

Appendices to "The Congestion Effect of Oil Transportation and Its Trade Implications"

C.1 The Presence of Search Frictions

In this appendix, we document one empirical fact pertaining to the presence of search frictions in the oil transportation industry, recognizing that such presence has been tested formally in Brancaccio et al. (2020). Using the shipping dataset, we find simultaneous arrivals of "empty and then load" ships, and departures of empty ships in many net exporting countries. Indeed, if we measure the under-utilization of empty ships by calculating the ratio, *outgoing empty ships/incoming empty and loading ships*, in the absence of search frictions, it should be close to zero (i.e., there are no unrealized matches when both parties are available). As seen in Figure C.1, however, even though the under-utilization ratio is often zero in Canada, it fluctuates significantly above zero in Saudi Arabia. In other words, it frequently happens that an empty ship arrives and picks up oil cargo, while at the same time another ship departs empty. In addition, if we calculate the average under-utilization ratio for the major net oil exporting countries, it is well above zero.

Figure C.1: Under-utilization of empty ships



Notes. The first two figures plot the under-utilization ratios for Canada and Saudi Arabia, respectively, during the sampling period from 2017W5 to 2020W48. The last figure plots the average under-utilization ratio for the first ten major net-oil-exporting countries in 2019, according to the JODI-Oil World Database.

C.2 Construction of the Weather Instruments

We construct the weather instruments in several steps. First, for each of the 15 major ports near the Persian Gulf (Fujairah, Jebelali, Jebel Dhanna Terminal, Khor Fakkan, Ruwais, and Vopak Terminal in the United Arab Emirates; Juaymah Terminal, King Fahd Port, and Ras Tanura in Saudi Arabia; Mina Abdulla and Mina Alahmadi in Kuwait; A1 Basra in Iraq and Ras Laffan in Qatar; Sitrah in Bahrain, and Sohar in Oman), we collect information on the horizontal wind speed at both 10m and 100m above the surface of the Earth, along with the air temperature at 2m above the surface of land, sea, or inland waters, from the ECMWF Reanalysis 5th Generation (ERA5) dataset, which provides hourly estimates of a large number of atmospheric, land, and oceanic climate variables. Next, we calculate the weekly averages of the weather variables at the 15 ports, with each port's data weighted by the average number of loaded trips originating from the country where the port is located from 2017-2020.

C.3 VAR Evidence of Shipping Market Congestion

Here we present additional evidence of shipping market congestion. Specifically, we follow Mercan et al. (2021) by constructing a vector autoregression (VAR) model

to study the response of the average contracting probability for ships in the oil transportation industry to an exogenous increase in the number of unloaded and searching ships (i.e., a tanker under-utilization shock). Specifically, we study the behaviors of three endogenous variables given by the vector:

$$\boldsymbol{y_{n,t}} = [\omega_{n,t}, \log s_{n,t}, \log f_{n,t}],$$

where $\omega_{n,t}$ is the measure of demand for oil transportation (following Kilian and Murphy (2014)), $s_{n,t}$ is the number of unloaded and searching ships (normalized by the total number of ships), and $f_{n,t}$ is the average contracting probability for ships in each market n. To construct $\omega_{n,t}$, we first develop a weekly measure of oil demand $x_{n,t}$ using the freight earnings data in the crude and product oil shipping markets, respectively. Specifically, we initialize $x_{n,2017W4} = 1$, and for each subsequent week through 2020W48, update $x_{n,t}$ using the average freight earnings index in either the crude oil market (BDTI Aframax TCE) or the product oil market (BCTI TC1-TCE). Both are deflated according to the U.S. Consumer Price Index:

$$x_{n,t} = \begin{cases} x_{n,t-1} + \Delta \log \text{BDTI}_t, & \text{if } n = \text{crude oil;} \\ x_{n,t-1} + \Delta \log \text{BCTI}_t, & \text{if } n = \text{product oil.} \end{cases}$$

We then regress $x_{n,t}$ on a linear time trend using heteroskedasticity-robust standard errors:

$$x_{n,t} = \beta_0 + \beta_1 t + \omega_{n,t},$$

and extract the residuals $\omega_{n,t}$ as the measure of demand for oil transportation services.

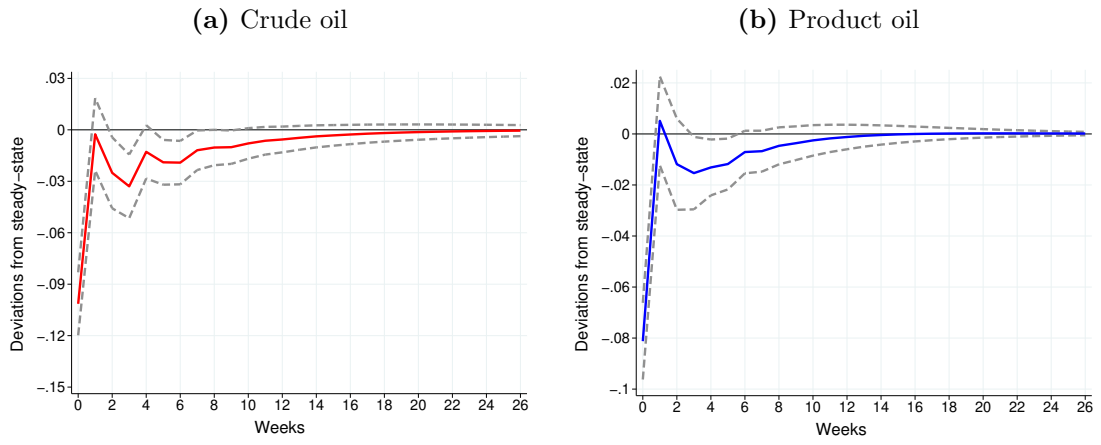
Subsequently, given $\boldsymbol{y_{n,t}}$, we estimate the following VAR model for each shipping market n:

$$\boldsymbol{y_{n,t}} = \boldsymbol{c_n} + \boldsymbol{A_n}(L)\boldsymbol{y_{n,t-1}} + \boldsymbol{u_{n,t}},$$

where $\boldsymbol{c_n}$ is a constant, $\boldsymbol{A_n}(L)$ is a lag polynomial, and $\boldsymbol{u_{n,t}} \sim (\boldsymbol{0}, \boldsymbol{\Sigma_u})$ is a vector of reduced-form error terms with variance-covariance matrix $\boldsymbol{\Sigma_u}$. We include three lags of the endogenous variables in our specification (based on the Akaike information criterion), and identify structural shocks using a recursive identification

scheme (or, equivalently, a Cholesky decomposition of $\mathbf{\Sigma}_{\boldsymbol{u}}$). We assume that ω has contemporaneous effects on both s and f. By contrast, s only has a contemporaneous effect on f, and f affects the other endogenous variables, only with a lag. Given the timing assumptions, we study the IRFs of the average contracting probability to an exogenous tanker under-utilization shock, so as to isolate the effect of movements in tanker under-utilization from those of demand fluctuations on the average contracting probability for ships.

Figure C.2: Empirical IRFs of average contracting probability to a tanker under-utilization shock



Notes. The above figures plot the IRFs of the average contracting probability for ships to a one standard deviation tanker under-utilization shock in each shipping market, using a three-variable VAR, identified off a Cholesky decomposition. The IRFs are plotted as solid lines, and the 95% confidence bands are plotted as dashed lines.

Figure C.2 plots the empirical impulse responses. The empirical responses are significantly negative on impact, indicating that shipping market congestion intensifies as exporters' capacity for hiring does not quickly and sufficiently expand to absorb the increasing number of searching ships. Subsequently, due to the imbalance between supply and demand for shipping services, exporters will endogenously respond by lowering freight rates, which in turn encourages hiring and boosts the contracting probability for oil tankers. Thus, we see a recovery in the average contracting probability for ships after impact, although it remains below the steady-state level while shipping market congestion lasts.

C.4 Derivation of the Integrated Value Functions

We begin with the value of a loaded ship in market n, $W_n(\mathbf{\Omega})$, as in Equation (4.7):

$$\begin{aligned} W_n(\mathbf{\Omega}) =& w_n(\mathbf{\Omega}) + (1-d_n)\beta \mathbb{E}_{\mathbf{\Omega}',\boldsymbol{\epsilon}'}(W_n(\mathbf{\Omega}') + \epsilon'_{n,i}) \\ &+ d_n \beta \mathbb{E}_{\mathbf{\Omega}',\boldsymbol{\epsilon}'}\Big\{ \max\Big[S_n(\mathbf{\Omega}') + \epsilon'_{n,i}, \max_{j\neq n\in N}(M_{nj}(\mathbf{\Omega}') + \epsilon'_{j,i})\Big]\Big\}. \end{aligned}$$

We integrate out the idiosyncratic taste shocks and simplify the value function using the definition of $\tilde{S}_{nj}$, which gives the following:

$$W_n(\mathbf{\Omega}) = w_n(\mathbf{\Omega}) + (1-d_n)\beta \mathbb{E}_{\mathbf{\Omega}'}(W_n(\mathbf{\Omega}')) + d_n \beta \mathbb{E}_{\mathbf{\Omega}'}\mathbb{E}_{\boldsymbol{\epsilon}'}\Big[\max_{j\in N}(\tilde{S}_{nj}(\mathbf{\Omega}') + \epsilon'_{j,i})\Big],$$

where $\mathbb{E}_{\mathbf{\Omega}'}$ is the expectation taken over the aggregate state of the shipping industry, and $\mathbb{E}_{\boldsymbol{\epsilon}'}$ is the expectation taken over the idiosyncratic taste shocks. Specifically, the last term, $\mathbb{E}_{\boldsymbol{\epsilon}'}\Big[\max_{j\in N}(\tilde{S}_{nj}(\mathbf{\Omega}') + \epsilon'_{j,i})\Big]$, can be calculated from $\rho\mathbb{E}_{\boldsymbol{\epsilon}'}\Big[\max_{j\in N}(\frac{\tilde{S}_{nj}(\mathbf{\Omega}')-\rho\gamma}{\rho} + \frac{\epsilon'_{j,i}+\rho\gamma}{\rho})\Big]$, and $\Big\{\frac{\epsilon'_{j,i}+\rho\gamma}{\rho}\Big\}_{j\in N} \sim Gumbel(0,1)$. Thus, if we define $X \equiv \max_{j\in N}\Big(\frac{\tilde{S}_{nj}(\mathbf{\Omega}')-\rho\gamma}{\rho} + \frac{\epsilon'_{j,i}+\rho\gamma}{\rho}\Big)$, the independence of $\Big\{\frac{\epsilon'_{j,i}+\rho\gamma}{\rho}\Big\}_{j\in N}$ implies that, for all real x, $Pr(X \leq x)$ is the product of the individual chances, $Pr(\frac{\tilde{S}_{nj}(\mathbf{\Omega}')-\rho\gamma}{\rho} + \frac{\epsilon'_{j,i}+\rho\gamma}{\rho} \leq x), \forall j \in N$. Taking logs and applying the basic properties of exponential functions yields the following:

$$\begin{aligned} \log Pr(X \leq x) &= \log \prod_{j\in N} Pr(\frac{\tilde{S}_{nj}(\mathbf{\Omega}') - \rho\gamma}{\rho} + \frac{\epsilon'_{j,i} + \rho\gamma}{\rho} \leq x) \\ &= \log \prod_{j\in N} \exp\Big[-\exp\Big(-x + \frac{\tilde{S}_{nj}(\mathbf{\Omega}') - \rho\gamma}{\rho}\Big)\Big] \\ &= -\sum_{j\in N} \exp\Big(-x + \frac{\tilde{S}_{nj}(\mathbf{\Omega}') - \rho\gamma}{\rho}\Big) \\ &= -\exp(-x)\Big[\sum_{j\in N} \exp\Big(\frac{\tilde{S}_{nj}(\mathbf{\Omega}') - \rho\gamma}{\rho}\Big)\Big] \\ &= -\exp\Big\{-x + \log\Big[\sum_{j\in N} \exp\Big(\frac{\tilde{S}_{nj}(\mathbf{\Omega}') - \rho\gamma}{\rho}\Big)\Big]\Big\}. \end{aligned}$$

Hence, we can derive that:

$$\begin{aligned} Pr(X \leq x) &= \exp\Big(-\exp\Big\{-x + \log\Big[\sum_{j\in N} \exp\Big(\frac{\tilde{S}_{nj}(\mathbf{\Omega}') - \rho\gamma}{\rho}\Big)\Big]\Big\}\Big) \\ &= \exp\Big(-\exp\Big\{-\frac{x - \log\Big[\sum_{j\in N} \exp\Big(\frac{\tilde{S}_{nj}(\mathbf{\Omega}')-\rho\gamma}{\rho}\Big)\Big]}{1}\Big\}\Big), \end{aligned}$$

which is exactly the cumulative distribution function (CDF) of a *Gumbel* distribution, with a location parameter equal to $\log\big[\sum_{j\in N}\exp\big(\frac{\tilde{S}_{nj}(\mathbf{\Omega}')-\rho\gamma}{\rho}\big)\big]$, and a scale parameter equal to 1, i.e., $X \equiv \max_{j\in N}\big(\frac{\tilde{S}_{nj}(\mathbf{\Omega}')-\rho\gamma}{\rho}+\frac{\epsilon'_{j,i}+\rho\gamma}{\rho}\big)\sim Gumbel\Big(\log\big[\sum_{j\in N}\exp\big(\frac{\tilde{S}_{nj}(\mathbf{\Omega}')-\rho\gamma}{\rho}\big)\big],1\Big)$. Given this, we can derive the mean of X in the following manner:

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\epsilon}'}(X) &= \mathbb{E}_{\boldsymbol{\epsilon}'}\big[\max_{j\in N}\big(\frac{\tilde{S}_{nj}(\mathbf{\Omega}')-\rho\gamma}{\rho}+\frac{\epsilon'_{j,i}+\rho\gamma}{\rho}\big)\big]\\ &= \gamma+\log\big[\sum_{j\in N}\exp\big(\frac{\tilde{S}_{nj}(\mathbf{\Omega}')-\rho\gamma}{\rho}\big)\big]\\ &= \gamma+\log\Big\{\exp(-\gamma)\big[\sum_{j\in N}\exp\big(\frac{\tilde{S}_{nj}(\mathbf{\Omega}')}{\rho}\big)\big]\Big\}\\ &= \log\big[\sum_{j\in N}\exp\big(\frac{\tilde{S}_{nj}(\mathbf{\Omega}')}{\rho}\big)\big],\end{aligned}$$

and therefore,

$$\mathbb{E}_{\boldsymbol{\epsilon}'}\big[\max_{j\in N}(\tilde{S}_{nj}(\mathbf{\Omega}')+\epsilon'_{j,i})\big]=\rho\log\big[\sum_{j\in N}\exp\big(\frac{\tilde{S}_{nj}(\mathbf{\Omega}')}{\rho}\big)\big].$$

Taking into account this derivation and rearranging the terms gives:

$$W_n(\mathbf{\Omega})=w_n(\mathbf{\Omega})+(1-d_n)\beta\mathbb{E}_{\mathbf{\Omega}'}(W_n(\mathbf{\Omega}'))+d_n\beta\rho\mathbb{E}_{\mathbf{\Omega}'}\Big\{\log\big[\sum_{j\in N}\exp(\frac{\tilde{S}_{nj}(\mathbf{\Omega}')}{\rho})\big]\Big\},$$

Similarly, using the result we just derived, we can rewrite the value functions for staying and ballasting ships, respectively, as:

$$S_n(\mathbf{\Omega})=-c^w+f_n(\theta_n)\beta\mathbb{E}_{\mathbf{\Omega}'}(W_n(\mathbf{\Omega}'))+(1-f_n(\theta_n))\beta\rho\mathbb{E}_{\mathbf{\Omega}'}\Big\{\log\big[\sum_{j\in N}\exp(\frac{\tilde{S}_{nj}(\mathbf{\Omega}')}{\rho})\big]\Big\},$$

$$M_{nj}(\mathbf{\Omega})=-c^s+(1-d_{nj})\beta\mathbb{E}_{\mathbf{\Omega}'}(M_{nj}(\mathbf{\Omega}'))+d_{nj}\beta\rho\mathbb{E}_{\mathbf{\Omega}'}\Big\{\log\big[\sum_{k\in N}\exp(\frac{\tilde{S}_{jk}(\mathbf{\Omega}')}{\rho})\big]\Big\}.$$

C.5 Derivation of Commissions

First, we can derive the ship's surplus, $W_n - S_n$, as follows:

$$W_n(\mathbf{\Omega})-S_n(\mathbf{\Omega})=w_n(\mathbf{\Omega})+c^w+(1-d_n-f_n(\theta_n))\beta\mathbb{E}_{\mathbf{\Omega}'}\Big\{W_n(\mathbf{\Omega}')-\rho\log\big[\sum_{j\in N}\exp(\frac{\tilde{S}_{nj}(\mathbf{\Omega}')}{\rho})\big]\Big\}.$$

Applying the surplus sharing rule $\eta(J_n(\mathbf{\Omega}) - V_n(\mathbf{\Omega})) = (1-\eta)(W_n(\mathbf{\Omega}) - S_n(\mathbf{\Omega}))$, and substituting in for J_n, yields:

$$\eta(p_n r_n - w_n(\mathbf{\Omega})) + \eta(1-d_n)\beta\mathbb{E}_{\mathbf{\Omega}'}(J_n(\mathbf{\Omega}') - V_n(\mathbf{\Omega}')) + \eta(\beta\mathbb{E}_{\mathbf{\Omega}'}(V_n(\mathbf{\Omega}')) - V_n(\mathbf{\Omega})) =$$
$$(1-\eta)(w_n(\mathbf{\Omega}) + c^w) + (1-\eta)(1-d_n - f_n(\theta_n))\beta\mathbb{E}_{\mathbf{\Omega}'}\Big\{W_n(\mathbf{\Omega}') - \rho\log\Big[\sum_{j\in N}\exp(\frac{\tilde{S}_{nj}(\mathbf{\Omega}')}{\rho})\Big]\Big\}.$$

Solving for w_n gives the following:

$$w_n(\mathbf{\Omega}) = \eta p_n r_n - (1-\eta)c^w + \eta(1-d_n)\beta\mathbb{E}_{\mathbf{\Omega}'}(J_n(\mathbf{\Omega}') - V_n(\mathbf{\Omega}')) + \eta(\beta\mathbb{E}_{\mathbf{\Omega}'}(V_n(\mathbf{\Omega}')) - V_n(\mathbf{\Omega}))$$
$$- (1-\eta)(1-d_n - f_n(\theta_n))\beta\mathbb{E}_{\mathbf{\Omega}'}\Big\{W_n(\mathbf{\Omega}') - \rho\log\Big[\sum_{j\in N}\exp(\frac{\tilde{S}_{nj}(\mathbf{\Omega}')}{\rho})\Big]\Big\}.$$

We then apply the surplus sharing rule for the next period, i.e., $\eta(J_n(\mathbf{\Omega}') - V_n(\mathbf{\Omega}')) = (1-\eta)(W_n(\mathbf{\Omega}') - S_n(\mathbf{\Omega}'))$, together with subtracting and adding $(1-\eta)(1-d_n - f_n(\theta_n))\beta\mathbb{E}_{\mathbf{\Omega}'}(S_n(\mathbf{\Omega}'))$:

$$w_n(\mathbf{\Omega}) = \eta p_n r_n - (1-\eta)c^w + (1-\eta)(1-d_n)\beta\mathbb{E}_{\mathbf{\Omega}'}(W_n(\mathbf{\Omega}') - S_n(\mathbf{\Omega}')) + \eta(\beta\mathbb{E}_{\mathbf{\Omega}'}(V_n(\mathbf{\Omega}')) - V_n(\mathbf{\Omega}))$$
$$- (1-\eta)(1-d_n - f_n(\theta_n))\beta\mathbb{E}_{\mathbf{\Omega}'}\Big\{W_n(\mathbf{\Omega}') - S_n(\mathbf{\Omega}') + S_n(\mathbf{\Omega}') - \rho\log\Big[\sum_{j\in N}\exp(\frac{\tilde{S}_{nj}(\mathbf{\Omega}')}{\rho})\Big]\Big\}$$

Collecting terms reduces the above equation to:

$$w_n(\mathbf{\Omega}) = \eta p_n r_n - (1-\eta)c^w + (1-\eta)f_n(\theta_n)\beta\mathbb{E}_{\mathbf{\Omega}'}(W_n(\mathbf{\Omega}') - S_n(\mathbf{\Omega}')) + \eta(\beta\mathbb{E}_{\mathbf{\Omega}'}(V_n(\mathbf{\Omega}')) - V_n(\mathbf{\Omega}))$$
$$- (1-\eta)(1-d_n - f_n(\theta_n))\beta\mathbb{E}_{\mathbf{\Omega}'}\Big\{S_n(\mathbf{\Omega}') - \rho\log\Big[\sum_{j\in N}\exp(\frac{\tilde{S}_{nj}(\mathbf{\Omega}')}{\rho})\Big]\Big\}.$$

By applying the surplus sharing rule for the next period again, and given that $f_n(\theta_n) = q_n(\theta_n)\theta_n$, we derive the following:

$$w_n(\mathbf{\Omega}) = \eta p_n r_n - (1-\eta)c^w + \eta q_n(\theta_n)\theta_n\beta\mathbb{E}_{\mathbf{\Omega}'}(J_n(\mathbf{\Omega}') - V_n(\mathbf{\Omega}')) + \eta(\beta\mathbb{E}_{\mathbf{\Omega}'}(V_n(\mathbf{\Omega}')) - V_n(\mathbf{\Omega}))$$
$$- (1-\eta)(1-d_n - f_n(\theta_n))\beta\mathbb{E}_{\mathbf{\Omega}'}\Big\{S_n(\mathbf{\Omega}') - \rho\log\Big[\sum_{j\in N}\exp(\frac{\tilde{S}_{nj}(\mathbf{\Omega}')}{\rho})\Big]\Big\}.$$

Now, from the value function for unmatched exporters in Equation (4.12), we derive that:

$$q_n(\theta_n)\theta_n\beta\mathbb{E}_{\mathbf{\Omega}'}(J_n(\mathbf{\Omega}') - V_n(\mathbf{\Omega}')) = \frac{\theta_n}{1-\delta}V_n(\mathbf{\Omega}) - \theta_n\beta\mathbb{E}_{\mathbf{\Omega}'}(V_n(\mathbf{\Omega}')).$$

Combining the expression with the commissions equation derived in the previous step yields:

$$w_n(\mathbf{\Omega}) = \eta p_n r_n - (1-\eta)c^w - \eta\Big[(1 - \frac{\theta_n}{1-\delta})V_n(\mathbf{\Omega}) - (1-\theta_n)\beta\mathbb{E}_{\mathbf{\Omega}'}(V_n(\mathbf{\Omega}'))\Big]$$
$$- (1-\eta)(1-d_n - f_n(\theta_n))\beta\mathbb{E}_{\mathbf{\Omega}'}\Big\{S_n(\mathbf{\Omega}') - \rho\log\Big[\sum_{j\in N}\exp(\frac{\tilde{S}_{nj}(\mathbf{\Omega}')}{\rho})\Big]\Big\},$$

which is equivalent to:

$$w_n(\mathbf{\Omega}) = \eta p_n r_n - \eta\Big[(1-\frac{\theta_n}{1-\delta})V_n(\mathbf{\Omega}) - (1-\theta_n)\beta\mathbb{E}_{\mathbf{\Omega}'}(V_n(\mathbf{\Omega}'))\Big] - (1-\eta)\Big(c^w + (1-d_n-f_n(\theta_n))\beta\rho\mathbb{E}_{\mathbf{\Omega}'}\Big\{\log\Big[\frac{\exp(S_n(\mathbf{\Omega}')/\rho)}{\sum_{j\in N}\exp(\tilde{S}_{nj}(\mathbf{\Omega}')/\rho)}\Big]\Big\}\Big).$$

Since the staying probability in Equation (4.10) equals $\pi_{nn}(\mathbf{\Omega}') = 1/\{\sum_{j\in N}\exp[(\tilde{S}_{nj}(\mathbf{\Omega}') - S_n(\mathbf{\Omega}'))/\rho]\}$, we further simplify the commissions equation as follows:

$$w_n(\mathbf{\Omega}) = \eta p_n r_n - \eta\Big[(1-\frac{\theta_n}{1-\delta})V_n(\mathbf{\Omega}) - (1-\theta_n)\beta\mathbb{E}_{\mathbf{\Omega}'}(V_n(\mathbf{\Omega}'))\Big] - (1-\eta)\Big\{c^w + (1-d_n-f_n(\theta_n))\beta\rho\mathbb{E}_{\mathbf{\Omega}'}\Big[\log\Big(1-\sum_{j\neq n\in N}\pi_{nj}(\mathbf{\Omega}')\Big)\Big]\Big\},$$

provided that $\pi_{nn}(\mathbf{\Omega}') = 1 - \sum_{j\neq n\in N}\pi_{nj}(\mathbf{\Omega}')$.

C.6 Steady-State Existence

Proposition 3. *Suppose that the Gumbel shocks $\boldsymbol{\epsilon}$ and $\boldsymbol{\epsilon}^e$ have full support, $\mathfrak{z}$ and $\mathscr{E}_n$ are finite, and $v_n \leq \frac{\mathscr{E}_n}{\delta}$, $\forall n \in N$, then there exists a steady-state allocation of ships and exporters over commodity shipping markets, $\{e_n^*, s_n^*, m_{nj}^*, v_n^*, \forall n, j \in N\}$.*

Proof. We first derive the steady-state shipping market stocks and flows. From Equation (4.18), the steady-state has the following form:

$$m_{nj} = \frac{\pi_{nj}}{d_{nj}}[(1-f_n)s_n + d_n e_n + \sum_{k\neq n\in N} d_{kn}m_{kn}], \tag{C.1}$$

where f_n represents the steady-state contracting probability for ships in market n. Similarly, the transition equation for stayers in Equation (4.17) is written as follows at the steady-state:

$$s_n = \pi_{nn}[(1-f_n)s_n + d_n e_n + \sum_{j\neq n\in N} d_{jn}m_{jn}]. \tag{C.2}$$

Substituting in for s_n and re-writing Equation (C.1) yields:

$$m_{nj} = \frac{\pi_{nj}}{d_{nj}\pi_{nn}}s_n. \tag{C.3}$$

Furthermore, re-arranging Equation (4.16) gives the steady-state stock of loaded ships:

$$e_n = \frac{f_n}{d_n}s_n. \tag{C.4}$$

Replacing m_{jn} and e_n in Equation (C.2) with expressions (C.3) and (C.4), respectively, we can re-write Equation (C.2) as follows:

$$s_n = \pi_{nn} s_n + \pi_{nn} \sum_{j \neq n \in N} \frac{\pi_{jn}}{\pi_{jj}} s_j, \tag{C.5}$$

which links the steady-state stock of stayers in market n, s_n, to those in the other markets, $\{s_j\}_{j \neq n \in N}$.

As for the stock of unmatched exporters described in Equation (4.19), its steady-state transition equation is given by:

$$v_n = (1 - \delta)(v_n - f_n s_n) + \mathscr{E}_n \pi_n^e, \tag{C.6}$$

which links the steady-state stock of unmatched exporters in market n, v_n, to the stock of stayers, s_n.

The steady-state transition equations (C.5) and (C.6) have a fixed point over a properly defined subset of $\mathscr{R}_+^{2N}$, according to Brouwer's Fixed Point Theorem (Fuente, 2000), which states that let $g : \mathscr{X} \to \mathscr{X}$ be a continuous function mapping a compact and convex subset of $\mathscr{R}_+^{2N}$, $\mathscr{X}$, into itself.[1] Then g has a fixed point in $\mathscr{X}$; that is, there exists at least one $x^* \in \mathscr{X}$ such that $g(x^*) = x^*$. Define $g : \mathscr{R}_+^{2N} \to \mathscr{R}_+^{2N}, g = (g^s, g^v)$ with:

$$\begin{aligned} g_n^s &= \pi_{nn} s_n + \pi_{nn} \sum_{j \neq n \in N} \frac{\pi_{jn}}{\pi_{jj}} s_j, \\ g_n^v &= (1 - \delta)(v_n - f_n s_n) + \mathscr{E}_n \pi_n^e, \end{aligned}$$

for all $n \in N$. Let $X \equiv \Theta_s \times \Theta_v$, where $\Theta_s \equiv \{s_n : \forall n \in N, s_n \geq 0, \sum_{n \in N} s_n \leq \mathscr{s}\}$, and $\Theta_v \equiv \{v_n : \forall n \in N, v_n \in [0, \frac{\mathscr{E}_n}{\delta}]\}$. X is non-empty, compact, and convex, while g is continuous on X. Now, we need to prove that g maps X into itself, i.e., $g(X) \subseteq X$. Let $\{s_n, v_n\}_{n \in N} \in X$. This implies that $\sum_{n \in N} s_n \leq \mathscr{s}$ and $v_n \leq \frac{\mathscr{E}_n}{\delta}$. Summing g_n^s over n yields:

$$\sum_{n \in N} g_n^s = \sum_{n \in N} \pi_{nn} s_n + \sum_{n \in N} (\pi_{nn} \sum_{j \neq n \in N} \frac{\pi_{jn}}{\pi_{jj}} s_j)$$

[1] Here we implicitly assume that the number of commodity varieties, N, is finite. We can easily generalize our results to infinite varieties, however, following the Leray-Schauder-Tychonoff Theorem.

$$= \sum_{n \in N} \pi_{nn} s_n + \sum_{n \in N} (1 - \pi_{nn}) s_n$$
$$= \sum_{n \in N} s_n \leq \mathcal{S}.$$

Hence, $g_n^s \in \Theta_s$. In terms of g_n^v, we can deduce that:

$$g_n^v \leq (1-\delta)v_n + \mathscr{E}_n \pi_n^e \leq (1-\delta)v_n + \mathscr{E}_n \leq (1-\delta)\frac{\mathscr{E}_n}{\delta} + \mathscr{E}_n = \frac{\mathscr{E}_n}{\delta}.$$

Therefore, $g_n^v \in \Theta_v$, and there exists at least one allocation of $\{s_n^*, v_n^*\}_{n \in N} \in X$ such that the steady-state transition equations (C.5) and (C.6) hold. As shown in Equations (C.3) and (C.4), given that $\{m_{nj}^*, e_n^*\}_{n,j \in N}$ are functions of $\{s_n^*\}_{n \in N} \in X$, there exists a steady-state allocation of ships and exporters over commodity shipping markets. □

C.7 Converting Model Moments to Data Moments

As discussed in Section 4.3.3.2, the moments generated by the model are not theoretically equivalent to what we observe in the shipping data, mainly due to the fact that the ballasting state is unobserved in reality. Given this complexity, we convert the model moments to their empirical counterparts, which we can retrospectively observe in the shipping data. To do so, we first expand the state space. In its original setup, the model includes three states of ships in each shipping market, namely, loaded $\{e_c, e_p\}$, unloaded and staying in the same market as before $\{s_c, s_p\}$, and unloaded and ballasting to a new market $\{m_{cp}, m_{pc}\}$. In order to keep track of each ship's operating market of last contract, we construct the following ten states:

$$\{e_c, e_p, s_c^c, s_c^p, s_p^c, s_p^p, m_{cp}^c, m_{cp}^p, m_{pc}^c, m_{pc}^p\},$$

with the superscript denoting a ship's market of last contract. Next, we formulate the transition matrix $P^{t,t+1}$ in the time-inhomogeneous Markov chain implied by the expansion of the states. For $t > 1$, the transition probability $P_{ij}^{t,t+1} \equiv Pr(state_{t+1} = j | state_t = i)$ is given by the i^{th} row and j^{th} column element of $P^{t,t+1}$:

$$P^{t,t+1} = \begin{pmatrix} 1-d_c & 0 & d_c\pi_{cc,t} & 0 & 0 & 0 & d_c\pi_{cp,t} & 0 & 0 & 0 \\ 0 & 1-d_p & 0 & 0 & 0 & d_p\pi_{pp,t} & 0 & 0 & 0 & d_p\pi_{pc,t} \\ f_{c,t} & 0 & (1-f_{c,t})\pi_{cc,t} & 0 & 0 & 0 & (1-f_{c,t})\pi_{cp,t} & 0 & 0 & 0 \\ f_{c,t} & 0 & 0 & (1-f_{c,t})\pi_{cc,t} & 0 & 0 & 0 & (1-f_{c,t})\pi_{cp,t} & 0 & 0 \\ 0 & f_{p,t} & 0 & 0 & (1-f_{p,t})\pi_{pp,t} & 0 & 0 & 0 & (1-f_{p,t})\pi_{pc,t} & 0 \\ 0 & f_{p,t} & 0 & 0 & 0 & (1-f_{p,t})\pi_{pp,t} & 0 & 0 & 0 & (1-f_{p,t})\pi_{pc,t} \\ 0 & 0 & 0 & 0 & d_{cp}\pi_{pp,t} & 0 & 1-d_{cp} & 0 & d_{cp}\pi_{pc,t} & 0 \\ 0 & 0 & 0 & 0 & 0 & d_{cp}\pi_{pp,t} & 0 & 1-d_{cp} & 0 & d_{cp}\pi_{pc,t} \\ 0 & 0 & d_{pc}\pi_{cc,t} & 0 & 0 & 0 & d_{pc}\pi_{cp,t} & 0 & 1-d_{pc} & 0 \\ 0 & 0 & 0 & d_{pc}\pi_{cc,t} & 0 & 0 & 0 & d_{pc}\pi_{cp,t} & 0 & 1-d_{pc} \end{pmatrix},$$

where $\pi_{nj,t}, \forall n, j \in \{c, p\}$ represent the staying and ballasting probabilities for ships, and $f_{c,t}, f_{p,t}$ represent the contracting probabilities for ships in the crude and product oil shipping markets, respectively. By evaluating the transition matrix at its steady-state, we study its stationary distribution, and pin down the implied steady-state value for each of the ten states. Specifically, we calculate the left eigenvector that corresponds to a unit eigenvalue, which then contains the steady-state number of ships in each state. Lastly, we apply these steady-state values as the initial allocation of ships in the Markov chain and compute the allocations thereafter.

The above method allows us to achieve consistency between the data and model-generated moments. For instance, the data-equivalent stock of unloaded ships in the crude oil shipping market can be written as $s_c^c + s_p^c + m_{cp}^c + m_{pc}^c$, which is essentially the total number of unloaded ships last contracted to carry crude oil. In addition, we derive the same notion of ballasting ships that is applied in the shipping data, i.e., ships that find oil cargoes to deliver in the market that differs from the previous one.

C.8 MCMC Approach to SMM Estimation

In this appendix, we describe the SMM procedures that are followed to estimate the eight free parameters in the baseline model, i.e., $\{\bar{r}, \mu_c, \psi_c, \psi_p, \rho, d_{cp}, \tau, \chi\}$. Let $\hat{\mathcal{S}}_T \in \mathscr{R}^8$ denote the set of data moments that will be matched in the estimation, with T denoting the sample size. Furthermore, let $\tilde{\mathcal{S}}(x)$ denote the moments implied by the model for some parameter value x. The SMM parameter estimate $\hat{x}_T$ is defined as:

$$\hat{x}_T = \arg\min_x \mathscr{L}(x),$$

where

$$\mathscr{L}(x) \equiv [\hat{\mathscr{S}}_T - \tilde{\mathscr{S}}(x)]' \hat{\Sigma}_{\mathscr{S},T}^{-1} [\hat{\mathscr{S}}_T - \tilde{\mathscr{S}}(x)],$$

in which $\hat{\Sigma}_{\mathscr{S},T}$ is an estimate of the variance-covariance matrix for the data moments $\hat{\mathscr{S}}_T$. The SMM estimate $\hat{x}_T$ chooses the model parameters, such that the model moments $\tilde{\mathscr{S}}(x)$ fit the observed data moments $\hat{\mathscr{S}}_T$ as closely as possible in terms of a quadratic form with weighting matrix $\hat{\Sigma}_{\mathscr{S},T}^{-1}$. We estimate $\hat{\Sigma}_{\mathscr{S},T}$ from the shipping data, using the Newey-West estimator, with a Bartlett kernel and bandwidth given by the integer of $4(T/100)^{2/9}$ (Ruge-Murcia, 2012). Note that the Newey-West weighting matrix does not depend on the model parameters or the model moments. Adapting standard results from the SMM procedures, one can prove that, for a given set of moments included in $\hat{\mathscr{S}}_T$, the estimate $\hat{x}_T$ is consistent and is the best estimate among those obtained with different weighting matrices.

In practice, the extremum estimator $\hat{x}_T$ is difficult to work with, since for any finite simulation length, the objective function is not smooth due to simulation errors and the discrete jumps that occur as a result of searching. To address this issue, we use the Markov Chain Monte Carlo (MCMC) approach for classical estimators proposed by Chernozhukov and Hong (2003) (which is further elaborated in Lise (2013)). Estimation proceeds by simulating a chain of parameters that (once converged) has the quasi-posterior density:

$$p(x) = \frac{\exp(\mathscr{L}(x))\pi(x)}{\int \exp(\mathscr{L}(x))\pi(x)dx},$$

with $\pi(x)$ representing the prior distribution of x. A point estimate for the parameters is obtained as the average of the MCMC chain, i.e., for $\{x^j, 0 \leq j \leq C\}$,

$$\hat{x}_{MCMC} = \frac{1}{C} \sum_{j=1}^{C} x^j.$$

In order to simulate a Markov chain that converges to the quasi-posterior, we apply the Metropolis-Hastings algorithm (Fernández-Villaverde et al., 2016). The algorithm generates a chain $\{x^j, 0 \leq j \leq C\}$ in the following steps:

1. Choose a starting value x^0;

2. Generate φ from a proposed density $q(\varphi|x^j)$;

3. Update x^{j+1} from x^j for $j = 1, 2, \ldots$ using:

$$x^{j+1} = \begin{cases} \varphi, & \text{with probability } d(x^j, \varphi); \\ x^j, & \text{with probability } 1 - d(x^j, \varphi), \end{cases}$$

where

$$d(x^j, \varphi) = \min[\frac{\exp(\mathscr{L}(x^j))\pi(x^j)q(\varphi|x^j)}{\exp(\mathscr{L}(\varphi))\pi(\varphi)q(x^j|\varphi)}, 1].$$

These procedures are repeated many times to obtain a chain of length C that represents the ergodic distribution of x.[2] Choosing the prior $\pi(x)$ to be a uniform distribution and the proposal density to be a random walk (i.e., $q(\varphi|x^j) = q(x^j|\varphi)$) results in the simple rule:

$$d(x^j, \varphi) = \min[\exp(\mathscr{L}(x^j) - \mathscr{L}(\varphi)), 1].$$

The main advantage of this estimation strategy is that it only requires function evaluations, and thus the discontinuous jumps do not cause the same problems that would otherwise occur with a gradient based extremum estimator. The drawback of this strategy is that it requires a very long chain, and consequently a very large number of function evaluations – each requiring the model to be solved and simulated. In our setup, we simulate 20 chains in parallel, each of length 2,000, and use the last 20 elements (pooled over the 20 chains) to obtain the parameter estimates in Table 4.3.

C.9 Recovering Data-Consistent Trade Shocks

To recover the market-specific trade shocks that are consistent with the observed evolution of tanker utilization from 2017W5 to 2020W48, we apply a first order approximation to the baseline model described in Section 4.3. In the model there are 37 endogenous variables, of which ten are state variables. Let $E_{37\times1}$ and $S_{10\times1}$ denote the endogenous and state variables, respectively. Then, the first

[2] We impose an additional restriction on the newly-generated φ, such that a steady-state of the model can be solved. This step is imposed to ensure only sensible draws of φ are rolled in the Markov chain.

order approximation of the baseline model solves for a linear relationship between E, S, and the innovations ν:

$$E_{t+1} - \bar{E} = A(S_t - \bar{S}) + B\nu_t,$$

where ν_t is the 2×1 vector of market-specific trade shocks, and A, B contain the transition coefficients solved by Dynare. To recover ν_t, we take the following steps:

1. Assume in period 0 that the economy is in a steady-state, hence $S_0 - \bar{S} = 0$. We treat 2017W4 as the steady-state;

2. Using the data on tanker utilization in the crude and product oil shipping markets in 2017W5 (Period 1), the first order approximation implies that there is exactly one solution for ν_1 that produces the observed tanker utilization in Period 1. We then solve for the implied $\hat{\nu}_1$;

3. After solving for $\hat{\nu}_1$, we solve for the remaining endogenous variables in E to recover $\hat{E}_1$ and $\hat{S}_1$;

4. We repeat steps 2 and 3 until the end of the tanker utilization series.

C.10 Baseline vs. Counterfactual Exercises

In Table C.1, we present the extended comparison between the baseline and the five counterfactual scenarios, including full utilization, no ballasting, zero trade variation, idling, and free-entry. The counterfactual exercises of idling and free-entry are discussed in detail in Appendices C.11 and C.12, respectively. The first and second moments of tanker utilization, oil trade, oil prices, and final good production are shown in the table.

C.11 Counterfactual: Idling of Ships

In this appendix we present another micro behavior on the part of ships – idling. The term idling refers to the state in which ships are temporarily idle due to a lack of cargo, or phased out of commercial operations. Ships often turn idle amid

Table C.1: Extended comparison: Baseline vs. counterfactual exercises

Scenario	e_c	e_p	y_c	y_p	p_c	p_p	Y
Baseline	0.200	0.432	0.224	0.487	0.843	1.058	0.704
	(0.016)	(0.023)	(12.387)	(15.071)	(27.352)	(12.481)	(12.924)
Counterfactual exercises							
Full utilization	0.313	0.687	0.328	0.753	0.433	0.997	0.989
	(0.000)	(0.000)	(19.798)	(24.084)	(0.847)	(0.828)	(22.873)
No ballasting	0.222	0.437	0.248	0.492	0.815	1.071	0.729
	(0.015)	(0.019)	(13.912)	(15.218)	(25.945)	(13.166)	(13.306)
Zero trade variation	0.200	0.432	0.201	0.433	0.842	1.059	0.627
	(0.016)	(0.023)	(0.016)	(0.023)	(0.031)	(0.014)	(0.026)
Idling	0.173	0.383	0.194	0.431	0.850	1.055	0.619
	(0.017)	(0.027)	(10.653)	(13.239)	(27.627)	(12.340)	(11.308)
Free-entry	0.205	0.461	0.229	0.519	0.856	1.052	0.741
	(0.054)	(0.074)	(12.603)	(16.055)	(28.060)	(12.231)	(13.687)

Notes. The above table shows the first and second moments of tanker utilization, the oil trade, oil prices, and final good production for each of the six scenarios: baseline, full utilization, no ballasting, zero trade variation, idling, and free-entry. $\{e_c, e_p\}$ refer to tanker utilization in the crude and product oil shipping markets, $\{y_c, y_p\}$ to the crude and product oil trades, $\{p_c, p_p\}$ to crude and product oil prices, and Y to final good production. The standard deviations are given in parentheses.

significant trade disruptions, since freight rates are not sufficient to cover their running costs and idling is usually preferred to the scrapping of the ship, considering how costly the ship-building process is.[3]

Despite its empirical relevance, idling is not taken into account in the baseline model for two reasons. First, according to Clarksons Shipping Intelligence Network, idle ships only represent a small proportion of the LR1 and LR2 oil tankers considered in our sample, at an average of 5% during the sampling period. The rareness is primarily due to the huge re-commissioning cost the shipowner incurs when bringing the ship back into commercial operations. Second, we cannot distinguish each individual ship's (potential) idling state from its unloaded state in the constructed shipping data, since unloaded ships resting in port may show the same observed behaviors. Furthermore, even though we could add the average share of idle ships

[3]See Kalouptsidi (2014) for a detailed study on ship-building.

in the fleet as an additional data moment in the SMM procedures, the other data moments (in particular, the transition probabilities) would be contaminated, since they do not take into account the transitions in and out of the idling state. For instance, if we were to consider the idling state explicitly, the contracting probability for ships should be adjusted to reflect that it is the group of *unloaded but not idle* ships that engage in searching for oil cargoes to deliver.

Nevertheless, we set up an idling counterfactual so as to study and quantify the changes that idling causes in the macro aggregates. Specifically, when ships are allowed to switch to the idling state, their value functions must be rewritten. Firstly, the value of a loaded ship i in market n (net of idiosyncratic taste shocks) is given by:

$$\begin{aligned} W_n(\mathbf{\Omega}) =& w_n(\mathbf{\Omega}) + (1-d_n)\beta\mathbb{E}_{\mathbf{\Omega}',\boldsymbol{\epsilon}'}(W_n(\mathbf{\Omega}') + \epsilon'_{n,i}) \\ &+ d_n\beta\mathbb{E}_{\mathbf{\Omega}',\boldsymbol{\epsilon}'}\Big\{ \max\Big[S_n(\mathbf{\Omega}') + \epsilon'_{n,i}, \max_{j\neq n\in N}(M_{nj}(\mathbf{\Omega}') + \epsilon'_{j,i}), L_n(\mathbf{\Omega}') + \epsilon'_{0,i}\Big]\Big\}, \end{aligned}$$

where $\mathbf{\Omega} = \{e_n, s_n, m_{nj}, l_n, v_n, r_n, \forall n, j \in N\}$ and $\epsilon'_{0,i}$ represents ship i's taste draw for idling (denoted by 0) for the next period. As embedded in $W_n(\mathbf{\Omega})$, once the ship fulfills its contract with probability d_n, it can choose whether to stay in market n, ballast from market n to some other market j, or become idle. In a similar fashion, the value of a staying ship in market n (net of idiosyncratic taste shocks) changes to:

$$\begin{aligned} S_n(\mathbf{\Omega}) =& -c^w + f_n(\theta_n)\beta\mathbb{E}_{\mathbf{\Omega}',\boldsymbol{\epsilon}'}(W_n(\mathbf{\Omega}') + \epsilon'_{n,i}) \\ &+ (1-f_n(\theta_n))\beta\mathbb{E}_{\mathbf{\Omega}',\boldsymbol{\epsilon}'}\Big\{ \max\Big[S_n(\mathbf{\Omega}') + \epsilon'_{n,i}, \max_{j\neq n\in N}(M_{nj}(\mathbf{\Omega}') + \epsilon'_{j,i}), L_n(\mathbf{\Omega}') + \epsilon'_{0,i}\Big]\Big\}, \end{aligned}$$

while the value of a mover ballasting from market n to j (net of idiosyncratic taste shocks) becomes:

$$\begin{aligned} M_{nj}(\mathbf{\Omega}) =& -c^s + (1-d_{nj})\beta\mathbb{E}_{\mathbf{\Omega}',\boldsymbol{\epsilon}'}(M_{nj}(\mathbf{\Omega}') + \epsilon'_{j,i}) \\ &+ d_{nj}\beta\mathbb{E}_{\mathbf{\Omega}',\boldsymbol{\epsilon}'}\Big\{ \max\Big[S_j(\mathbf{\Omega}') + \epsilon'_{j,i}, \max_{k\neq j\in N}(M_{jk}(\mathbf{\Omega}') + \epsilon'_{k,i}), L_j(\mathbf{\Omega}') + \epsilon'_{0,i}\Big]\Big\}. \end{aligned}$$

Lastly, we define the value of an idle ship in market n (net of idiosyncratic taste shocks) as:

$$L_n(\mathbf{\Omega}) = -c^w + \beta\mathbb{E}_{\mathbf{\Omega}',\boldsymbol{\epsilon}'}\Big\{ \max\Big[S_n(\mathbf{\Omega}') + \epsilon'_{n,i}, \max_{j\neq n\in N}(M_{nj}(\mathbf{\Omega}') + \epsilon'_{j,i}), L_n(\mathbf{\Omega}') + \epsilon'_{0,i}\Big]\Big\},$$

which states that the idle ship pays the per-period waiting cost c^w, and is free to choose between re-entering the market and remaining idle in the next period.

Following the changes in the value functions, we rewrite the transition equations for shipping market stock (including the stock of idle ships), and build an alternative model of idling using the same set of parameters as described in Table 4.3. Again, we apply a first order approximation to the model and solve for the implied shipping market stocks and macro aggregates using the same set of recovered trade shocks as in all the other scenarios. We present the means and standard deviations of tanker utilization, the oil trade, oil prices, and final good production in Table C.1.

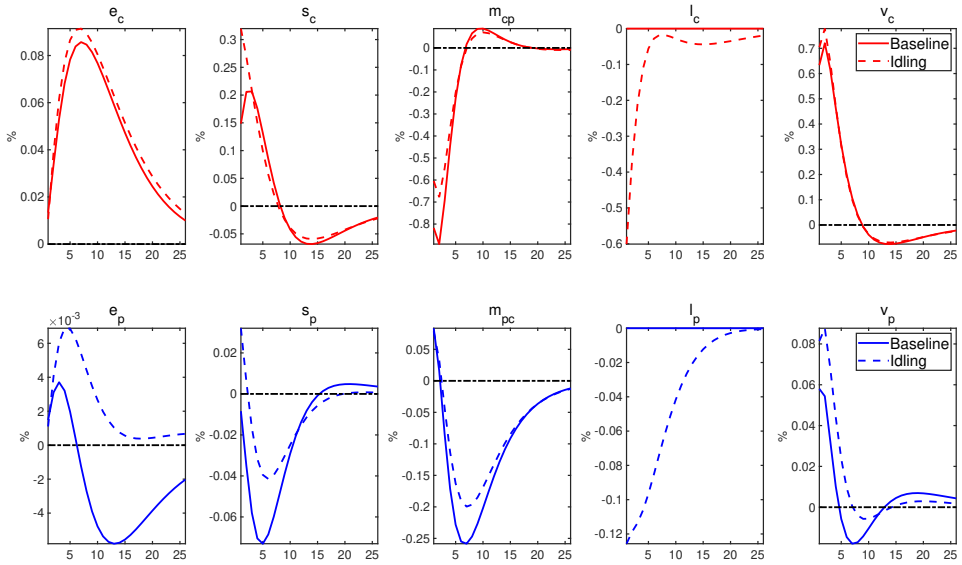
What happens to the average level of the oil trade? Solving for the implied shipping market stock yields an average share of idle ships of 2.5% in the crude oil market and 7.3% in the product oil market. This occurs, on the one hand, because ships can now shield themselves from adverse trade shocks by not only ballasting to the other market, but also by becoming idle. Furthermore, in cases in which both shipping markets are hit by a *global* adverse trade shock, ships might even find themselves in a better position by turning idle rather than ballasting to the other market. On the other hand, when the shipping market receives a favorable shock, since there are always gross flows into the idle state, fewer ships will be operating to carry oil.[4] In consequence, as shown in Table C.1, irrespective of the type of trade shocks, the average oil trade decreases in the idling counterfactual, with crude and product oil trade seeing declines of 13.4% and 11.5%, respectively, against the baseline.

What happens to the volatility of the oil trade? In Figure C.3, we plot the IRFs of shipping market stock to an orthogonal, positive crude oil trade shock in the idling counterfactual. By comparing them with those corresponding to the baseline scenario, we first observe that the stock of idle ships in the crude oil market l_c plummets significantly on impact, implying that more ships "awake" and actively participate in searching again. Their transitions out of the idle state increase the stock of stayers s_c, hence intensifying the congestion for ships but weakening that for exporters. As a result, m_{cp} declines relatively less, while v_c jumps marginally

[4]This is because the idling probability is always positive in this alternative model, i.e., $\pi_{n0}(\mathbf{\Omega}') = 1/\sum_{j\in N}\exp\left[(\tilde{S}_{nj}(\mathbf{\Omega}') - \tilde{S}_{n0}(\mathbf{\Omega}'))/\rho\right] > 0$, where $\tilde{S}_{n0}(\mathbf{\Omega}') = L_n(\mathbf{\Omega}')$.

more on impact. The combined effect is that tanker utilization e_c reacts more significantly to the positive trade shock, and so does the crude oil trade. In the product oil market, the substitution effect still plays its role in inducing more potential exporters to enter the market and take advantage of the higher product oil price. We note, however, that v_p rises more significantly than in the baseline, because they no longer need to compete for a shrinking supply of transportation services: Since idle ships are intrigued by the higher commissions and the ease of finding oil to deliver, more of them become active and hence s_p increases on impact. Consequently, the initial rise in e_p is more dramatic, thus amplifying the fluctuations in the product oil trade. In this sense, compared to the baseline, enabling idling destabilizes the oil trade, not only in the market in which a trade shock hits, but also in the one where it does not. This result also holds in the case of a positive product oil trade shock, for which the IRFs are plotted in Figure C.8.[5]

Figure C.3: IRFs to a positive crude oil trade shock: Baseline vs. idling



Notes. The IRFs plot the dynamic weekly responses of shipping market stock (i.e., $\{e_{n,t}, s_{n,t}, m_{nj,t}, l_{n,t}, v_{n,t}, \forall n, j \in \{c, p\}\}$) to one standard deviation of an orthogonal trade shock in the crude oil shipping market ν_c for the baseline and idling counterfactual, respectively. The shock ν_c is identical in both scenarios. The IRFs are computed using a second order approximation of each alternative model, and are measured in percentage deviations from the steady-state.

[5]Note that when idling is enabled, a trade shock no longer induces the inverse movements of tanker utilization in the crude and product oil shipping markets (see bottom left panels in Figures C.3 and C.8). This is because the outflow of ships from the idle state can replenish the shrinking pool of stayers in the market where the shock does not hit.

Lastly, although the fluctuations in the oil trade are exacerbated in the process of shipping markets endogenously responding to external shocks, the lower average oil trade resulting from idling ultimately stabilizes the oil trade: As illustrated in Table C.1, the standard deviations of the crude and product oil trades decline by 14.0% and 12.2%, respectively.

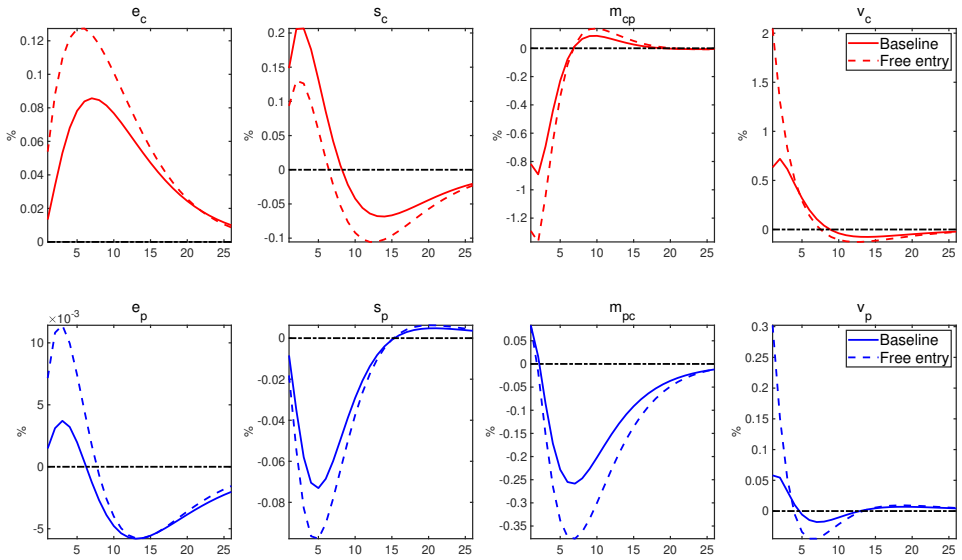
C.12 Counterfactual: Free-Entry of Potential Exporters

In the last counterfactual exercise, we take a look at the potential exporter's cost of entry, and its impacts on shipping market congestion and oil trade fluctuations. In doing so, we consider a free-entry counterfactual in which potential exporters can enter the shipping markets freely and search for available ships to deliver oil cargoes. In the model, this implies that the value of an exporter that remains unmatched in shipping market n, V_n, becomes zero, and thus the free-entry condition is written as:

$$\frac{\kappa}{\beta q_n(\theta_n)} = \mathbb{E}_{\mathbf{\Omega}'}(J_n(\mathbf{\Omega}')),$$

where κ stands for the per-period searching cost paid by an exporter while looking for a ship. It replaces Equation (4.19), and determines the stock of oil cargoes to be delivered in each shipping market. To solve this counterfactual model, we again follow the parameterization in Table 4.3 and set the searching cost κ at unity. In particular, the calibration of κ is based on the idea that the assumed free-entry condition permits one normalization, as in Shimer (2005): Doubling κ and multiplying the market-specific matching efficiency ψ_n by $2^{1-\alpha}$ doubles $q_n(\theta_n)$ for each market, yet leaves $f_n(\theta_n)$ unchanged. Once again, after applying a first order approximation to the counterfactual model, we solve for the implied shipping market stocks and macro aggregates using the same set of recovered trade shocks as in the other scenarios. We present the corresponding means and standard deviations of tanker utilization, the oil trade, oil prices, and final good production in the last row of Table C.1.

Figure C.4: IRFs to a positive crude oil trade shock: Baseline vs. free-entry



Notes. The IRFs plot the dynamic weekly responses of shipping market stocks (i.e., $\{e_{n,t}, s_{n,t}, m_{nj,t}, v_{n,t}, \forall n, j \in \{c, p\}\}$) to one standard deviation of an orthogonal trade shock in the crude oil shipping market ν_c for the baseline and free-entry counterfactual, respectively. The shock ν_c is identical in both scenarios. The IRFs are computed using a second order approximation of each alternative model, and are measured in percentage deviations from the steady-state.

In order to differentiate the model's dynamics in the free-entry counterfactual from those embedded in the baseline scenario, we again study the IRFs of the shipping market stock, which are shown in Figures C.4 and C.9. We focus on the case of a positive trade shock in the crude oil shipping market. When the shock hits, the stock of crude oil to be delivered shoots up more significantly because of free-entry. The more responsive reaction of v_c raises commissions and the contracting probability for ships by a wider margin, resulting in fewer ships ballasting to the product oil market and significantly higher tanker utilization in the crude oil market. In the product oil market, since the substitution effect between crude and product oil is stronger due to the larger increase in e_c, we see a much greater on-impact response of tanker utilization, e_p. Consequently, both the crude and product oil trade rises and fluctuates more.

C.13 Natural Experiments

Through the counterfactual exercises, we demonstrate the vital role that oil transportation plays in the propagation of oil trade shocks to the wider economy and its significant implications for macro aggregates. To complement these findings and explore how changes in the model's primitives affect the micro behaviors of ships and exporters and subsequently give rise to macro consequences, we consider three natural experiments: shipping fuel regulations to cut sulfur emissions (IMO 2020), the 2021 Suez Canal obstruction, and the possibility of an unsanctioned Iran. These three experiments highlight the inner workings of the different parts of the baseline model – from the ballasting behavior of ships to the entry of potential exporters into a shipping market.

C.13.1 IMO 2020

The first natural experiment is related to the shipping fuel regulations on vessel sulfur emissions enacted by the International Maritime Organization (IMO), effective from 1 January 2020. The new regulations, known as IMO 2020, mandate a maximum sulfur content of 0.5% in marine fuel globally in order to reduce air pollution created by the shipping industry. Although some transportation research (e.g., Van et al. (2019) and references therein) has studied the implications of sulfur regulation for public health, the environment, and the oil refinery process, there has been no attempt to explore its implications for shipping market composition, the oil trade, and final good production.

In order to shed light on this question, we apply our estimated framework by considering a hypothetical scenario in which sailing costs double, since the regulation directly affects the cost of transportation and thus changes the optimal behavior of ships. Table C.2 presents several outcomes of interest. Perhaps contrary to conventional wisdom (i.e., higher sailing costs hurt exporters' profit margins, hence reducing the oil trade and final good production), we find that tanker utilization actually increases in both markets by 0.68% and 0.61%, respectively. As a result, both the crude and product oil trades, together with final good production, rise by

Table C.2: Baseline vs. natural experiments

Moments	Baseline	(1) IMO 2020	(1) % Change	(2) Suez Blockage	(2) % Change	(3) Unsanctioned Iran	(3) % Change
Shipping markets							
Tanker utilization							
Crude oil	0.196	0.198	0.679%	0.200	2.054%	0.196	0.042%
Product oil	0.430	0.432	0.607%	0.447	3.984%	0.430	0.073%
Staying ships							
Crude oil	0.105	0.104	-1.153%	0.098	-7.082%	0.105	-0.140%
Product oil	0.232	0.233	0.245%	0.225	-3.260%	0.232	-0.054%
Ballasting ships							
Crude oil → product oil	0.015	0.014	-8.362%	0.013	-16.269%	0.015	-0.317%
Product oil → crude oil	0.021	0.019	-9.693%	0.017	-17.412%	0.021	-0.354%
Oil to be delivered							
Crude oil	0.046	0.046	1.352%	0.045	-0.954%	0.046	0.144%
Product oil	0.143	0.145	1.115%	0.144	0.481%	0.143	0.169%
Commissions							
Crude oil	0.585	0.583	-0.347%	0.599	2.293%	0.585	0.042%
Product oil	0.671	0.669	-0.292%	0.681	1.482%	0.671	0.028%
Oil & final good markets							
Crude oil price	0.846	0.846	-0.026%	0.852	0.689%	0.846	0.011%
Product oil price	1.057	1.057	0.009%	1.054	-0.250%	1.057	-0.004%
Crude oil trade	0.197	0.198	0.679%	0.201	2.054%	0.197	0.042%
Product oil trade	0.431	0.434	0.607%	0.449	3.984%	0.432	0.073%
Final good production	0.622	0.626	0.626%	0.644	3.466%	0.622	0.064%

Notes. The moments in each experiment are calculated as the averages of 10,000 simulations, based on a second order approximation of the baseline model. The % change for each variable of interest in each experiment is calculated against the baseline for comparison. The IMO 2020 is modeled as a natural experiment with a twofold increase in the ship's sailing cost, i.e., $c^s = -0.024, c^w = -0.025$. The 2021 Suez Canal obstruction is modeled as a natural experiment with a 5% decrease in the ship's arrival probability within each shipping market, i.e., $d_c = 0.0438, d_p = 0.0353$. The possibility of an unsanctioned Iran is modeled as a natural experiment with a 5% increase in the stock of potential exporters in each shipping market, i.e., $\mathscr{E}_c = 1.0815, \mathscr{E}_p = 0.8400$.

0.68%, 0.61%, and 0.63%, respectively. This increase in tanker utilization unveils an often overlooked channel through which the ships and exporters endogenously adjust their behaviors in response to a rise in transportation costs. Higher sailing costs reduce the outside option value for ships by lowering the value (and thus the probability) of ballasting. Since the commissions have this particular form:

$$\begin{aligned} w_n(\mathbf{\Omega}_n) =& \eta p_n r_n - \eta\Big[(1-\frac{\theta_n}{1-\delta})V_n(\theta_n) - (1-\theta_n)\beta\mathbb{E}_{\mathbf{\Omega}'}(V_n(\mathbf{\Omega}'))\Big] \\ & - (1-\eta)\Big\{c^w + (1-d_n-f_n(\theta_n))\beta\rho\mathbb{E}_{\mathbf{\Omega}'}\Big[\log\Big(1-\sum_{j\neq n\in N}\pi_{nj}(\mathbf{\Omega}')\Big)\Big]\Big\}, \end{aligned}$$

where $\pi_{nj}(\mathbf{\Omega}')$ is the ballasting probability for ships from market n to j, the lower ballasting probability π_{nj} suppresses commissions and encourages more cargoes to be delivered, resulting in increases in tanker utilization, the oil trade, and final good production. To put it differently, increased sailing costs imply that ships are more "tied" to their current jobs, and since their bargaining position is weaker, they negotiate lower commissions. Such predictions are confirmed in Table C.2 as (1) the stocks of ballasting ships decrease by 8.36% and 9.69%; (2) the oil cargoes to be delivered increase by 1.35% and 1.12%; and (3) commissions fall by 0.35% and 0.29% in the crude and product oil shipping markets, respectively.

C.13.2 2021 Suez Canal Obstruction

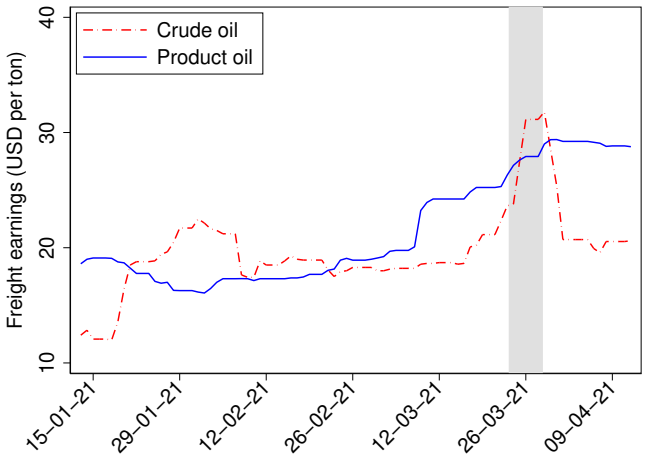
The second natural experiment is related to the 2021 Suez Canal blockage, which occurred from 23-29 March 2021 after strong winds and a dust storm caused the 200,000 DWT MV Ever Given (a Panama-flagged vessel operated by the Taiwanese shipping company Evergreen) to become wedged sideways across the canal. As a result, the blocked canal delayed about 30 fuel-carrying tankers with approximately 8.8 million barrels of crude oil, amounting to a little less than one tenth of a day's global consumption and worth about $550 million (Reed, 2021).

We estimate the impact of such an obstruction by modeling it as a 5% decrease in the arrival probability for ships within each shipping market. As a significant proportion of crude and product oil passes through the Suez Canal (see Figure 4.1), a blockage directly increases the loaded ships' traveling time and essentially makes them "stuck" in the loaded state and unable to ballast for better opportunities.[6] In consequence, they are compensated with higher commissions, which encourages tanker utilization in the shipping markets. From Table C.2, we can clearly see that commissions rose by 2.29% and 1.48% following the Suez Canal obstruction, and that tanker utilization roses by 2.05% and 3.98% in the crude and product oil shipping markets, respectively. Such predictions are externally validated, as

[6]Note that ballasting ships rarely take the Suez Canal, primarily because of its passage fee; they normally take an alternative route, passing by the Cape of Good Hope and Cape Agulhas at the southern edge of Africa.

we observe a sharp increase in freight earnings in the crude oil shipping market amid the Suez blockage, as shown in Figure C.5.

Figure C.5: Freight earnings amid the Suez Canal blockage



Notes. The plotted series are dollars per ton freight earnings in two major trading routes for crude and product oil, respectively (namely, BCTI-TC8 (Middle East Gulf to UK-Cont.) and BDTI-TD23 (Middle East Gulf to Mediterranean)), both of which involve passage through the Suez Canal. The shaded area represents the time window of the Suez Canal blockage, 23-29 March 2021. Data on the freight earnings come from the Baltic Exchange.

Triggered by the increases in commissions and tanker utilization, the crude and product oil trade rises significantly by 2.05% and 3.98%, respectively. Since the product oil trade increases proportionately more than the crude oil trade (as predicted by the downward sloping demand curve in Equation (4.21)), the product oil price declines marginally, by 0.25%, while the crude oil price jumps significantly, by 0.69%. Lastly, since the oil trade is projected to increase as a result of the Suez Canal blockage, final good production will rise accordingly.

C.13.3 An Unsanctioned Iran

In the final experiment, we consider the possibility of an unsanctioned Iran and its impact on the oil trade and final good production. As one of the largest oil producing and exporting countries in the world, Iran has been subject to strict U.S. sanctions since 2018, when the Trump administration unilaterally withdrew from the Iran nuclear deal that restricted the Middle Eastern country's nuclear activities, and subsequently ended exemptions from the U.S. secondary sanctions

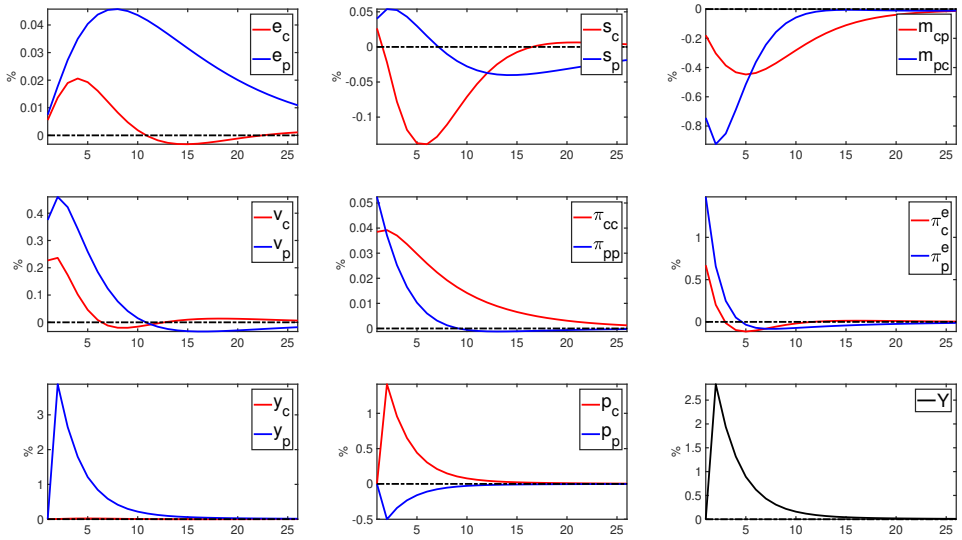
for major importers of Iranian oil. As a result, Iran's oil exports have plummeted from the peak of 2.5 million barrels per day in early 2018 to 0.3 million barrels per day in late 2019. Recently, President Joe Biden has been expected to seek the restoration of that accord, and officials in Tehran have expressed the hope that he will ease restrictions on their petroleum sales.[7] In order to simulate the effects of Iran's return to oil sales, we increase the number of potential exporters in the crude and product oil markets by 5%.[8] Table C.2 presents the resulting outcomes compared to the baseline simulations. Unsurprisingly, the oil cargoes to be delivered in the crude and product oil shipping markets increase by 0.14% and 0.17%, respectively. The resulting higher demand for oil transportation services translates to a greater payoff for ships to carry oil, with commissions rising by 0.04% and 0.03% in the crude and product oil shipping markets, respectively. In consequence, tanker utilization increases in both markets, leading to increases in the oil trade and final good production (though only marginally when compared to the rise in the stock of potential exporters).

C.14 Additional Impulse Response Functions

[7]For more detailed coverage of the Biden administration's intent to restore the 2015 nuclear agreement with Iran and to ease sanctions, see https://foreignpolicy.com/2021/04/15/iran-nuclear-deal-biden-talks-vienna/ (Accessed 22 May 2021).

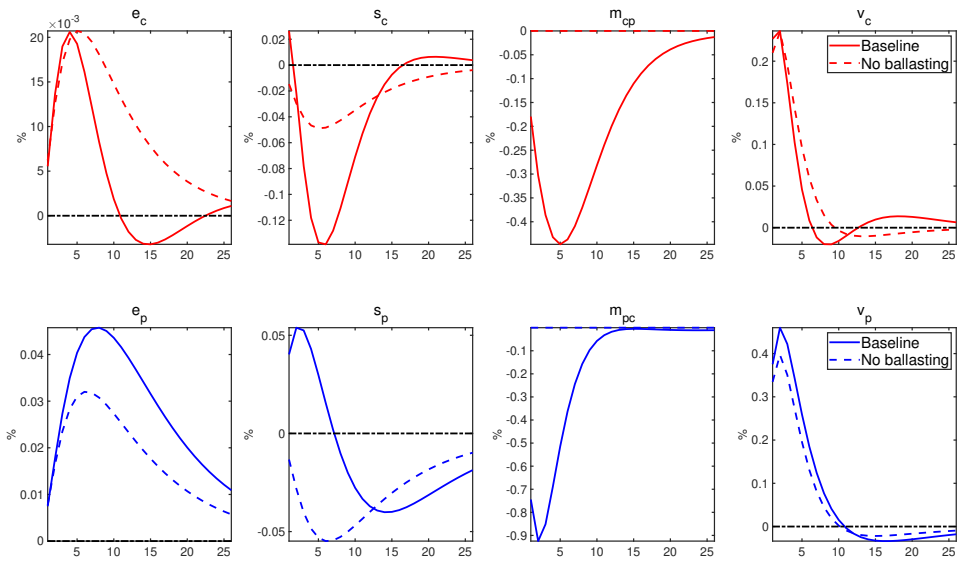
[8]We select 5% in accordance with the pre-sanction oil production level by Iran. According to the U.S. Energy Information Administration, before the sanctions, Iran accounted for approximately 4.4% of total daily oil production in the world, producing just over 4 million barrels per day in 2018.

Figure C.6: IRFs to a positive product oil trade shock



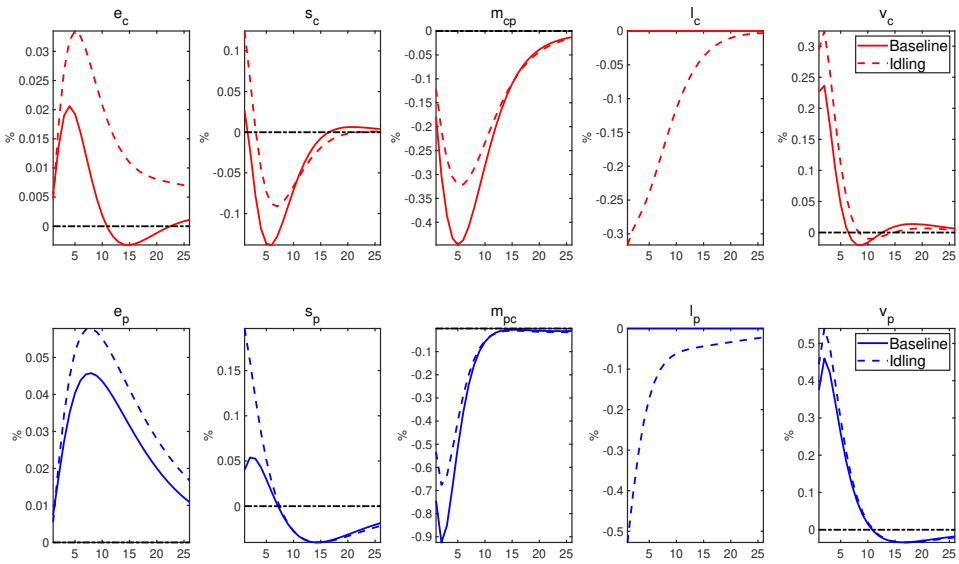
Notes. The IRFs plot the dynamic weekly responses of selected endogenous variables to one standard deviation of an orthogonal trade shock in the product oil shipping market ν_p. The IRFs are computed using a second order approximation of the baseline model, and are measured in percentage deviations from the steady-state.

Figure C.7: IRFs to a positive product oil trade shock: baseline vs. no ballasting



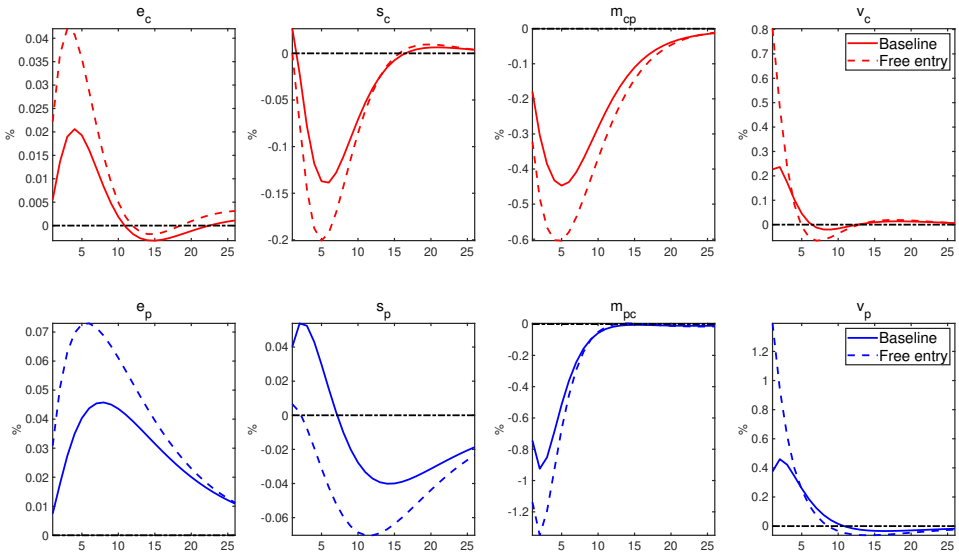
Notes. The IRFs plot the dynamic weekly responses of shipping market stocks (i.e., $\{e_{n,t}, s_{n,t}, m_{nj,t}, v_{n,t}, \forall n, j \in \{c,p\}\}$) to one standard deviation of an orthogonal trade shock in the product oil shipping market ν_p for the baseline and no ballasting counterfactual, respectively. The shock ν_p is identical in both scenarios. The IRFs are computed using a second order approximation of each alternative model, and are measured in percentage deviations from the steady-state.

Figure C.8: IRFs to a positive product oil trade shock: Baseline vs. idling



Notes. The IRFs plot the dynamic weekly responses of shipping market stocks (i.e., $\{e_{n,t}, s_{n,t}, m_{nj,t}, l_{n,t}, v_{n,t}, \forall n, j \in \{c, p\}\}$) to one standard deviation of an orthogonal trade shock in the product oil shipping market ν_p for the baseline and idling counterfactual, respectively. The shock ν_p is identical in both scenarios. The IRFs are computed using a second order approximation of each alternative model, and are measured in percentage deviations from the steady-state.

Figure C.9: IRFs to a positive product oil trade shock: Baseline vs. free-entry



Notes. The IRFs plot the dynamic weekly responses of shipping market stocks (i.e., $\{e_{n,t}, s_{n,t}, m_{nj,t}, v_{n,t}, \forall n, j \in \{c, p\}\}$) to one standard deviation of an orthogonal trade shock in the product oil shipping market ν_p for the baseline and free-entry counterfactual, respectively. The shock ν_p is identical in both scenarios. The IRFs are computed using a second order approximation of each alternative model, and are measured in percentage deviations from the steady-state.

References

Adamopoulos, T., & Restuccia, D. (2020). Land reform and productivity: A quantitative analysis with micro data. *American Economic Journal: Macroeconomics*, *12*(3), 1–39. https://doi.org/10.1257/mac.20150222

Adland, R., Cariou, P., & Wolff, F.-C. (2016). The influence of charterers and owners on bulk shipping freight rates. *Transportation Research Part E: Logistics and Transportation Review*, *86*, 69–82. https://doi.org/10.1016/j.tre.2015.11.014

Adland, R., Hansson, D., & Wense, L. v. d. (2017). Valuing cargo flexibility in oil transportation. *Maritime Policy & Management*, *44*(7), 803–814. https://doi.org/10.1080/03088839.2017.1366079

Adland, R., Jia, H., & Strandenes, S. P. (2017). Are AIS-based trade volume estimates reliable? The case of crude oil exports. *Maritime Policy & Management*, *44*(5), 657–665. https://doi.org/10.1080/03088839.2017.1309470

Alessandrini, A., Mazzarella, F., & Vespe, M. (2018). Estimated time of arrival using historical vessel tracking data. *IEEE Transactions on Intelligent Transportation Systems*, *20*(1), 7–15. https://doi.org/10.1109/TITS.2017.2789279

Allen, T. (2014). Information frictions in trade. *Econometrica*, *82*(6), 2041–2083. https://doi.org/10.3982/ECTA10984

Arslanalp, S., Marini, M., & Tumbarello, P. (2019). Big data on vessel traffic: Nowcasting trade flows in real time. *IMF Working Paper No. 2019/275*. https://www.imf.org/en/Publications/WP/Issues/2019/12/13/Big-Data-on-Vessel-Traffic-Nowcasting-Trade-Flows-in-Real-Time-48837

Artuç, E., Chaudhuri, S., & McLaren, J. (2010). Trade shocks and labor adjustment: A structural empirical approach. *American Economic Review*, *100*(3), 1008–45. https://doi.org/10.1257/aer.100.3.1008

Asturias, J. (2020). Endogenous transportation costs. *European Economic Review*, *123*, 103366. https://doi.org/10.1016/j.euroecorev.2019.103366

Bai, X., Hou, Y., & Yang, D. (2021). Choose clean energy or green technology? Empirical evidence from global ships. *Transportation Research Part E:*

Logistics and Transportation Review, *151*, 102364. https://doi.org/10.1016/j.tre.2021.102364

Bai, X., & Lam, J. S. L. (2019). A destination choice model for very large gas carriers (VLGC) loading from the US Gulf. *Energy*, *174*, 1267–1275. https://doi.org/10.1016/j.energy.2019.02.148

Bai, X., & Li, Y. (2021). The congestion effect of oil transportation and its trade implications. *Working Paper*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3851173

Balgova, M. (2020). Leaping into the unknown? The role of job search in migration decisions. *Working Paper*. https://mariabalgova.github.io/JMPBalgova.pdf

Balistreri, E. J., Al-Qahtani, A., & Dahl, C. A. (2010). Oil and petroleum product Armington elasticities: A new-geography-of-trade approach to estimation. *The Energy Journal*, *31*(3). http://www.jstor.org/stable/41323299

Beine, M., Docquier, F., & Rapoport, H. (2001). Brain drain and economic growth: theory and evidence. *Journal of Development Economics*, *64*(1), 275–289. https://doi.org/10.1016/S0304-3878(00)00133-4

Beine, M., Docquier, F., & Rapoport, H. (2008). Brain drain and human capital formation in developing countries: winners and losers. *The Economic Journal*, *118*(528), 631–652. https://doi.org/10.1111/j.1468-0297.2008.02135.x

Bhagwati, J., & Hamada, K. (1974). The brain drain, international integration of markets for professionals and unemployment: a theoretical analysis. *Journal of Development Economics*, *1*(1), 19–42. https://doi.org/10.1016/0304-3878(74)90020-0

Boddin, D., & Stähler, F. (2018). The organization of international trade. *CESifo Working Paper No. 7378*. https://ideas.repec.org/p/ces/ceswps/_7378.html

Bodenstein, M., Erceg, C. J., & Guerrieri, L. (2011). Oil shocks and external adjustment. *Journal of International Economics*, *83*, 168–184. https://doi.org/10.1016/J.JINTECO.2010.10.006

Bojnec, Š., & Dries, L. (2005). Causes of changes in agricultural employment in Slovenia: Evidence from micro-data. *Journal of Agricultural Economics*, *56*(3), 399–416. https://onlinelibrary.wiley.com/doi/10.1111/j.1477-9552.2005.00020.x

Bornstein, G., Krusell, P., & Rebelo, S. (2022). A world equilibrium model of the oil market. *The Review of Economic Studies*. https://doi.org/10.1093/RESTUD/RDAC019

Bosch, M. (2016). Does unemployment insurance offer incentives to take jobs in the formal sector? *IZA World of Labor 2016: 300*. https://doi.org/10.15185/izawol.300

Brancaccio, G., Kalouptsidi, M., & Papageorgiou, T. (2020). Geography, transportation, and endogenous trade costs. *Econometrica*, *88*(2), 657–691. https://doi.org/10.3982/ECTA15455

Brandt, M. W., & Gao, L. (2019). Macro fundamentals or geopolitical events? A textual analysis of news events for crude oil. *Journal of Empirical Finance*, *51*, 64–94. https://doi.org/10.1016/j.jempfin.2019.01.007

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Breithaupt, S. A., Copping, A., Tagestad, J., & Whiting, J. (2017). Maritime route delineation using AIS data from the Atlantic coast of the US. *The Journal of Navigation*, *70*(2), 379. https://doi.org/10.1017/S0373463316000606

Brenner, T., & Gildner, A. (2006). The long-term implications of local industrial clusters. *European Planning Studies*, *14*(9), 1315–1328. https://doi.org/10.1080/09654310600933371

Briones, R. M. (2017a). Characterization of agricultural workers in the Philippines. *PIDS Discussion Paper Series No. 2017-31*. https://pidswebs.pids.gov.ph/CDN/PUBLICATIONS/pidsdps1731.pdf

Briones, R. M. (2017b). Outlook for the Philippine economy and agro-industry to 2030: The role of productivity growth. *PIDS Discussion Paper Series No. 2017-30*. https://pidswebs.pids.gov.ph/CDN/PUBLICATIONS/pidsdps1730.pdf

Bryan, G., Chowdhury, S., & Mobarak, A. M. (2014). Underinvestment in a profitable technology: The case of seasonal migration in Bangladesh. *Econometrica*, *82*(5), 1671–1748. https://doi.org/10.3982/ECTA10489

Buchholz, N. (2021). Spatial equilibrium, search frictions, and dynamic efficiency in the taxi industry. *The Review of Economic Studies*. https://doi.org/10.1093/restud/rdab050

Bustos, P., Castro-Vincenzi, J., Monras, J., & Ponticelli, J. (2020). Industrialization without innovation. *NBER Working Paper No. 25871*. https://doi.org/10.3386/w25871

Caliendo, L., Dvorkin, M., & Parro, F. (2019). Trade and labor market dynamics: General equilibrium analysis of the china trade shock. *Econometrica*, *87*(3), 741–835. https://doi.org/10.3982/ECTA13758

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: methods and applications*. Cambridge University Press.

Camp, K. M., Mead, D., Reed, S. B., Sitter, C., & Wasilewski, D. (2020). From the barrel to the pump: the impact of the COVID-19 pandemic on prices for petroleum products. *Monthly Labor Review, U.S. Bureau of Labor Statistics*. https://doi.org/10.21916/mlr.2020.24

Carvalho, L. S. et al. (2010). Poverty and time preference. *RAND Labor and Population Working Paper Series No. 759*. https://doi.org/10.7249/WR759

Caselli, F. (2005). Accounting for cross-country income differences. *Handbook of Economic Growth*, *1*, 679–741. https://personal.lse.ac.uk/casellif/papers/handbook.pdf

Cerdeiro, D. A., Komaromi, A., Liu, Y., & Saeed, M. (2020). World seaborne trade in real time: A proof of concept for building AIS-based nowcasts from scratch. *IMF Working Paper No. 2020/57*. https://www.imf.org/en/Publications/WP/Issues/2020/05/14/World-Seaborne-Trade-in-Real-Time-A-Proof-of-Concept-for-Building-AIS-based-Nowcasts-from-49393

Chatziantoniou, I., Filippidis, M., Filis, G., & Gabauer, D. (2021). A closer look into the global determinants of oil price volatility. *Energy Economics*, *95*, 105092. https://doi.org/10.1016/j.eneco.2020.105092

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2011). SMOTE: Synthetic Minority Over-sampling Technique. *Journal Of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Chen, X. M., Zahiri, M., & Zhang, S. (2017). Understanding ridesplitting behavior of on-demand ride services: An ensemble learning approach. *Transportation Research Part C: Emerging Technologies*, *76*, 51–70. https://doi.org/10.1016/j.trc.2016.12.018

Chernozhukov, V., & Hong, H. (2003). An MCMC approach to classical estimation. *Journal of Econometrics*, *115*(2), 293–346. https://doi.org/10.1016/S0304-4076(03)00100-3

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, *6*(1), 3–73.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge. https://doi.org/10.4324/9780203774441

Coşar, A. K., & Demir, B. (2018). Shipping inside the box: Containerization and trade. *Journal of International Economics*, *114*, 331–345. https://doi.org/10.1016/J.JINTECO.2018.07.008

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, *2*(4), 303–314. https://doi.org/10.1007/BF02551274

Den Haan, W. J., Haefke, C., & Ramey, G. (2005). Turbulence and unemployment in a job matching model. *Journal of the European Economic Association*, *3*(6), 1360–1385. https://www.jstor.org/stable/40004938

Dennis, S. M. (2004). Improvements in BTS estimation of ton-miles. *U.S. Department of Transportation. Bureau of Transportation Statistics Working Paper No. 2004-002-OAS*. https://doi.org/https://doi.org/10.21949/1501502

Donaldson, D. (2018). Railroads of the Raj: Estimating the impact of transportation infrastructure. *American Economic Review*, *108*(4-5), 899–934. https://doi.org/10.1257/aer.20101199

Eaton, J., & Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, *70*(5), 1741–1779. https://www.jstor.org/stable/3082019

Elsby, M. W., Michaels, R., & Solon, G. (2009). The ins and outs of cyclical unemployment. *American Economic Journal: Macroeconomics*, *1*(1), 84–110. https://doi.org/10.1257/mac.1.1.84

Elsby, M. W., & Shapiro, M. D. (2012). Why does trend growth affect equilibrium employment? A new explanation of an old puzzle. *American Economic Review*, *102*(4), 1378–1413. https://doi.org/10.1257/aer.102.4.1378

European Commission. (2017). Modernizing and simplifying the CAP: Socio-Economic challenges facing agriculture and rural areas. https://ec.europa.eu/info/sites/default/files/food-farming-fisheries/key_policies/documents/soc_background_final_en.pdf

Fajgelbaum, P. D., & Schaal, E. (2020). Optimal transport networks in spatial equilibrium. *Econometrica*, *88*(4), 1411–1452. https://doi.org/10.3982/ECTA15213

Felipe, J., Bayudan-Dacuycuy, C., & Lanzafame, M. (2014). The declining share of agricultural employment in the People's Republic of China: How fast? *ADB Economics Working Paper Series No. 419*. https://www.adb.org/sites/default/files/publication/149676/ewp-419.pdf

Fernández-Villaverde, J., Rubio-Ramírez, J. F., & Schorfheide, F. (2016). Solution and estimation methods for DSGE models. *Handbook of macroeconomics* (pp. 527–724). Elsevier. https://doi.org/10.1016/bs.hesmac.2016.03.006

Flood, S., King, M., Rodgers, R., Ruggles, S., & Warren, J. R. (2020). Integrated public use microdata series, Current Population Survey: Version 7.0 [dataset] [Minneapolis, MN: IPUMS, 2020.]. https://doi.org/10.18128/D030.V7.0

Fox, J. T., Yang, C., & Hsu, D. H. (2018). Unobserved heterogeneity in matching games. *Journal of Political Economy*, *126*(4), 1339–1373. https://doi.org/10.1086/697740

Frechette, G. R., Lizzeri, A., & Salz, T. (2019). Frictions in a competitive, regulated market: Evidence from taxis. *American Economic Review*, *109*(8), 2954–92. https://doi.org/10.1257/aer.20161720

Fuente, A. d. l. (2000). *Mathematical methods and models for economists*. Cambridge University Press. https://doi.org/10.1017/CBO9780511810756

Fuglie, K. (2012). Productivity growth and technology capital in the global agricultural economy. *Productivity Growth In Agriculture: An International Perspective*, 335. https://doi.org/10.1079/9781845939212.0335

Fuglie, K. (2015). Accounting for growth in global agriculture. *Bio-Based and Applied Economics*, *4*(3), 201–234. https://doi.org/10.13128/BAE-17151

Fujita, S. (2018). Declining labor turnover and turbulence. *Journal of Monetary Economics*, *99*, 1–19. https://doi.org/10.1016/j.jmoneco.2017.12.005

Gervais, M., Jaimovich, N., Siu, H. E., & Yedid-Levi, Y. (2016). What should I be when I grow up? Occupations and unemployment over the life cycle. *Journal of Monetary Economics*, *83*, 54–70. https://doi.org/10.1016/j.jmoneco.2016.08.003

Gillis, C. (2020). US lifts sanctions against COSCO's tanker operation [*FreightWaves*, January 31, 2020. Accessed January 27, 2022.]. https://www.freightwaves.com/news/us-lifts-sanctions-against-coscos-tanker-operation

Gollin, D., Lagakos, D., & Waugh, M. E. (2014). The agricultural productivity gap. *The Quarterly Journal of Economics*, *129*(2), 939–993. https://doi.org/10.1093/qje/qjt056

Hamilton, T. M. (2014). Oil tanker sizes range from general purpose to ultra-large crude carriers on AFRA scale [*U.S. Energy Information Administration*, September 16, 2014. Accessed May 22, 2021.]. https://www.eia.gov/todayinenergy/detail.php?id=17991

Herrendorf, B., Rogerson, R., & Valentinyi, A. (2014). Growth and structural transformation. *Handbook of economic growth* (pp. 855–941). Elsevier. https://doi.org/10.1016/B978-0-444-53540-5.00006-9

Hosios, A. J. (1990). On the efficiency of matching and related models of search and unemployment. *The Review of Economic Studies*, *57*(2), 279–298. https://doi.org/10.2307/2297382

Hubbard, B., Karasz, P., & Reed, S. (2019). Two Major Saudi Oil Installations Hit by Drone Strike, and U.S. Blames Iran [*The New York Times*, September 15, 2019. Accessed October 5, 2021.]. https://www.nytimes.com/2019/09/14/world/middleeast/saudi-arabia-refineries-drone-attack.html

International Monetary Fund. (1999). Chronic unemployment in the Euro area: Causes and cures. https://fac.comtech.depaul.edu/jwoo1/courses/chronic-unemployment.pdf

Jia, H., Prakash, V., & Smith, T. (2019). Estimating vessel payloads in bulk shipping using AIS data. *International Journal of Shipping and Transport Logistics*, *11*(1), 25–40. https://doi.org/10.1504/IJSTL.2019.096864

Juhn, C., Murphy, K. M., & Topel, R. H. (2002). Current unemployment, historically contemplated. *Brookings Papers on Economic Activity*, *2002*(1), 79–116. https://doi.org/10.1353/eca.2002.0008

Juhn, C., Murphy, K. M., Topel, R. H., Yellen, J. L., & Baily, M. N. (1991). Why has the natural rate of unemployment increased over time? *Brookings Papers on Economic Activity*, *1991*(2), 75–142. https://doi.org/10.2307/2534590

Kalouptsidi, M. (2014). Time to build and fluctuations in bulk shipping. *American Economic Review*, *104*(2), 564–608. https://doi.org/10.1257/aer.104.2.564

Känzig, D. R. (2021). The macroeconomic effects of oil supply news: Evidence from opec announcements. *American Economic Review*, *111*(4), 1092–1125. https://doi.org/10.1257/aer.20190964

Kennan, J., & Walker, J. R. (2011). The effect of expected income on individual migration decisions. *Econometrica*, *79*(1), 211–251. https://doi.org/10.3982/ECTA4657

Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review*, *99*(3), 1053–69. https://doi.org/10.1257/aer.99.3.1053

Kilian, L., & Murphy, D. P. (2014). The role of inventories and speculative trading in the global market for crude oil. *Journal of Applied econometrics*, *29*(3), 454–478. https://doi.org/10.1002/jae.2322

Kilian, L., & Zhou, X. (2018). Modeling fluctuations in the global demand for commodities. *Journal of International Money and Finance*, *88*, 54–78. https://doi.org/10.1016/j.jimonfin.2018.07.001

Kimball, S. (2019). Trump says tariffs on $200 billion of Chinese goods will increase to 25%, blames slow progress in trade talks [*CNBC*, May 5, 2019. Accessed September 28, 2021.]. https://www.cnbc.com/2019/05/05/trump-says-tariffs-on-200-billion-of-chinese-goods-will-increase-to-25percent-on-friday.html

Kohn, D., Leibovici, F., & Tretvoll, H. (2021). Trade in commodities and business cycle volatility. *American Economic Journal: Macroeconomics*, *13*(3), 173–208. https://www.aeaweb.org/articles?id=10.1257/mac.20180131

Koirala, K. H., Mishra, A., & Mohanty, S. (2016). Impact of land ownership on productivity and efficiency of rice farmers: The case of the Philippines. *Land Use Policy*, *50*, 371–378. https://doi.org/10.1016/j.landusepol.2015.10.001

Kroodsma, D. A., Mayorga, J., Hochberg, T., Miller, N. A., Boerder, K., Ferretti, F., Wilson, A., Bergman, B., White, T. D., Block, B. A., et al. (2018). Tracking the global footprint of fisheries. *Science*, *359*(6378), 904–908. URL%20=%20%7Bhttps://www.science.org/doi/abs/10.1126/science.aao5646%7D

Lagakos, D., & Waugh, M. E. (2013). Selection, agriculture, and cross-country productivity differences. *American Economic Review*, *103*(2), 948–80. https://doi.org/10.1257/aer.103.2.948

Landler, M. (2018). Trump abandons Iran nuclear deal he long scorned [*The New York Times*, May 8, 2018. Accessed September 28, 2021.]. https://www.nytimes.com/2018/05/08/world/middleeast/trump-iran-nuclear-deal.html

Lin, Y., & Jeon, Y. (2006). Random forests and adaptive nearest neighbors. *Journal of the American Statistical Association*, *101*(474), 578–590. https://www.jstor.org/stable/27590719

Lise, J. (2013). On-the-job search and precautionary savings. *Review of Economic Studies*, *80*(3), 1086–1113. https://doi.org/10.1093/restud/rds042

Liu, L., Wang, Y., Wu, C., & Wu, W. (2016). Disentangling the determinants of real oil prices. *Energy Economics*, *56*, 363–373. https://doi.org/10.1016/j.eneco.2016.04.003

Liu, Z., Liu, Y., Meng, Q., & Cheng, Q. (2019). A tailored machine learning approach for urban transport network flow estimation. *Transportation Research Part C: Emerging Technologies*, *108*, 130–150. https://doi.org/10.1016/j.trc.2019.09.006

Llanto, G. M. (2012). The impact of infrastructure on agricultural productivity. *PIDS Discussion Paper Series No. 2012-12*. https://dirp4.pids.gov.ph/ris/dps/pidsdps1212.pdf

Matsuyama, K. (1992). A simple model of sectoral adjustment. *The Review of Economic Studies*, *59*(2), 375–387. https://doi.org/10.2307/2297959

Matzkin, R. L. (2003). Nonparametric estimation of nonadditive random functions. *Econometrica*, *71*(5), 1339–1375. https://www.jstor.org/stable/1555505

Maucorps, A., Münch, A., Brkanovic, S., Schuh, B., Dwyer, J. C., Vigani, M., Khafagy, A., Coto Sauras, M., Deschellette, P., Lopez, A., et al. (2019). The EU farming employment: Current challenges and future prospects. https://www.europarl.europa.eu/RegData/etudes/STUD/2019/629209/IPOL_STU(2019)629209_EN.pdf

McCaig, B., & Pavcnik, N. (2013). Moving out of agriculture: Structural change in Vietnam. *NBER Working Paper No. 19616*. https://doi.org/10.3386/w19616

McMillan, M. S., & Rodrik, D. (2011). Globalization, structural change and productivity growth. *NBER Working Paper No. 17143*. https://doi.org/10.3386/w17143

Melek, N. Ç., Plante, M., & Yücel, M. K. (2017). The US shale oil boom, the oil export ban, and the economy: A general equilibrium analysis. *NBER Working Paper No. 23818*. https://doi.org/10.3386/w23818

Mendoza, E. G. (1991). Real business cycles in a small open economy. *The American Economic Review*, 797–818. https://www.jstor.org/stable/2006643

Menzio, G., Telyukova, I. A., & Visschers, L. (2016). Directed search over the life cycle. *Review of Economic Dynamics*, *19*, 38–62. https://doi.org/10.1016/j.red.2015.03.002

Mercan, Y., Schoefer, B., & Sedláček, P. (2021). A Congestion Theory of Unemployment Fluctuations. *NBER Working Paper No. 28771*. https://cepr.org/active/publications/discussion_papers/dp.php?dpno=15500

Mortensen, D. T., & Pissarides, C. A. (1994). Job creation and job destruction in the theory of unemployment. *The Review of Economic Studies*, *61*(3), 397–415. https://doi.org/10.2307/2297896

Németh, G., Szabó, L., & Ciscar, J. C. (2011). Estimation of armington elasticities in a cge economy–energy–environment model for europe. *Economic Modelling*, *28*(4), 1993–1999. https://doi.org/10.1016/j.econmod.2011.03.032

Organisation for Economic Co-operation and Development. (1994). The OECD jobs study: Evidence and explanations. https://www.worldcat.org/title/oecd-jobs-study-evidence-and-explanations/oclc/32046198

Ostoj, I. (2015). Labor market efficiency as one of the pillars of the global competitiveness of an economy–conclusions for the labor market regimes of the EU countries. *Journal of Economics & Management*, *20*, 80–92. http:

//cejsh.icm.edu.pl/cejsh/element/bwmeta1.element.desklight-c1fb0fae-d1f8-49d2-a8aa-95bbd0e78e56

Petrongolo, B., & Pissarides, C. A. (2001). Looking into the black box: A survey of the matching function. *Journal of Economic Literature*, *39*(2), 390–431. https://doi.org/10.1257/jel.39.2.390

Pilossoph, L. (2014). Sectoral shocks and move unemployment. *Working Paper*.

Poten, & Partners. (2014). LR.2 Sides to the Story [*Poten Tanker Opinion*, December 19, 2014. Accessed September 14, 2021.]. https://www.poten.com/lr-2-sides-story/

Probst, P., & Boulesteix, A.-L. (2017). To tune or not to tune the number of trees in random forest? *Journal of Machine Learning Research*, *18*(1), 6673–6690. https://dl.acm.org/doi/10.5555/3122009.3242038

Probst, P., Wright, M. N., & Boulesteix, A.-L. (2019). Hyperparameters and tuning strategies for random forest. *WIREs Data Mining and Knowledge Discovery*, *9*(3), e1301. https://doi.org/https://doi.org/10.1002/widm.1301

Prochazka, V., Adland, R., & Wolff, F.-C. (2019). Contracting decisions in the crude oil transportation market: Evidence from fixtures matched with AIS data. *Transportation Research Part A: Policy and Practice*, *130*, 37–53. https://doi.org/10.1016/j.tra.2019.09.009

Qian, R., Warwick, M. K., Diop, N., Hansl, B., Cruz, K. T. G., Bruckner, M., Chua, K. C., Enriquez, K. A. L., Galang, R. M. N., Garcia, A. F., et al. (2018). Growth and productivity in the Philippines: Winning the future. http://hdl.handle.net/10986/30450

Raval, A., & Sheppard, D. (2017). Oil prices hit 2017 lows [*Financial Times*, May 4, 2017. Accessed March 9, 2022.]. https://www.ft.com/content/cb50714f-b213-3b9c-bf8b-f542bff0c7df

Reed, S. (2021). The oil market can absorb a blocked Suez Canal, for now [*The New York Times*, March 25, 2021. Accessed April 13, 2021.]. https://www.nytimes.com/2021/03/25/business/oil-prices-suez-canal.html

Regli, F. (2019). *Essays on crude oil tanker markets* (Doctoral dissertation). Denmark, Copenhagen Business School.

Regli, F., & Nomikos, N. K. (2019). The eye in the sky – Freight rate effects of tanker supply. *Transportation Research Part E: Logistics and Transportation Review*, *125*, 402–424. https://doi.org/10.1016/j.tre.2019.03.015

Restuccia, D., Yang, D. T., & Zhu, X. (2008). Agriculture and aggregate productivity: A quantitative cross-country analysis. *Journal of Monetary Economics*, *55*(2), 234–250. https://doi.org/10.1016/j.jmoneco.2007.11.006

Rizov, M., & Swinnen, J. F. (2004). Human capital, market imperfections, and labor reallocation in transition. *Journal of Comparative Economics*, *32*(4), 745–774. https://doi.org/10.1016/j.jce.2004.08.002

Rosegrant, M. W., Perez, N., Pradesha, A., & Thomas, T. S. (2016). The economywide impacts of climate change on Philippine agriculture. *IFPRI Policy Note 1*. https://www.ctc-n.org/sites/www.ctc-n.org/files/129755.pdf

Ruge-Murcia, F. (2012). Estimating nonlinear DSGE models by the simulated method of moments: With an application to business cycles. *Journal of Economic Dynamics and Control*, *36*(6), 914–938. https://doi.org/10.1016/j.jedc.2012.01.008

Scheyder, E., & Seba, E. (2017). Harvey throws a wrench into U.S. energy engine [*Reuters*, August 28, 2017. Accessed August 3, 2021.]. https://www.reuters.com/article/us-storm-harvey-energy-idUSKCN1B70YQ

Schwab, K. (2019). The global competitiveness report 2018. https://www.weforum.org/reports/the-global-competitveness-report-2018

Shimer, R. (2005). The cyclical behavior of equilibrium unemployment and vacancies. *American Economic Review*, *95*(1), 25–49. https://doi.org/10.1257/0002828053828572

Shimer, R. (2012). Reassessing the ins and outs of unemployment. *Review of Economic Dynamics*, *15*(2), 127–148. https://doi.org/10.1016/j.red.2012.02.001

Simonsen, M., Gössling, S., & Walnum, H. J. (2019). Cruise ship emissions in Norwegian waters: A geographical analysis. *Journal of Transport Geography*, *78*, 87–97. https://doi.org/10.1016/j.jtrangeo.2019.05.014

Sjaastad, L. A. (1962). The costs and returns of human migration. *Journal of Political Economy*, *70*(5, Part 2), 80–93. https://www.jstor.org/stable/1829105

Sødal, S., Koekebakker, S., & Aadland, R. (2008). Market switching in shipping – A real option model applied to the valuation of combination carriers. *Review of Financial Economics*, *17*(3), 183–203. https://doi.org/10.1016/j.rfe.2007.04.001

Statista. (2021). Transport volume of crude oil in global seaborne trade from 2010 to 2020 (in million metric tons) [*Statista: Water Transport*, May 20, 2021. Accessed March 13, 2022.]. https://www.statista.com/statistics/264013/transport-volume-of-crude-oil-in-seaborne-trade/

Stock, J., & Yogo, M. (2005). Testing for weak instruments in linear iv regression. *Identification and inference for econometric models* (pp. 80–108). Cambridge

University Press. https://scholar.harvard.edu/stock/publications/testing-weak-instruments-linear-iv-regression

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, *8*(1), 1–21. https://doi.org/10.1186/1471-2105-8-25

Sugrue, D., & Adriaens, P. (2021). A data fusion approach to predict shipping efficiency for bulk carriers. *Transportation Research Part E: Logistics and Transportation Review*, *149*, 102326. https://doi.org/10.1016/j.tre.2021.102326

Teruel, R. G., & Kuroda, Y. (2005). Public infrastructure and productivity growth in Philippine agriculture, 1974–2000. *Journal of Asian Economics*, *16*(3), 555–576. https://doi.org/10.1016/j.asieco.2005.04.011

Tocco, B., Bailey, A., & Davidova, S. (2013). Determinants to leave agriculture and change occupational sector: Evidence from an enlarged EU. *CEPS Factor Markets Working Paper Series No. 46*. https://www.ceps.eu/ceps-publications/determinants-leave-agriculture-and-change-occupational-sector-evidence-enlarged-eu/

Trafigura. (2019). *2019 Annual report* (tech. rep.). Trafigura Group Pte. Ltd.

United Nations Conference on Trade and Development. (2019). Review of Maritime Transport 2019. https://unctad.org/system/files/official-document/rmt2019_en.pdf

U.S. Energy Information Administration. (2021). Country analysis executive summary: Iran [*U.S. Energy Information Administration*, July 20, 2021. Accessed September 28, 2021.]. https://www.eia.gov/international/analysis/country/IRN

Van, T. C., Ramirez, J., Rainey, T., Ristovski, Z., & Brown, R. J. (2019). Global impacts of recent IMO regulations on marine fuel oil refining processes and ship emissions. *Transportation Research Part D: Transport and Environment*, *70*, 123–134. https://doi.org/10.1016/j.trd.2019.04.001

Winther, M., Christensen, J. H., Plejdrup, M. S., Ravn, E. S., Eriksson, Ó. F., & Kristensen, H. O. (2014). Emission inventories for ships in the arctic based on satellite sampled AIS data. *Atmospheric Environment*, *91*, 1–14. https://doi.org/10.1016/j.atmosenv.2014.03.006

Wong, W. F. (2018). The round trip effect: Endogenous transport costs and international trade. *Working Paper*. https://www.freit.org/WorkingPapers/Papers/Transportation/FREIT1361.pdf

Xia, Y., Kong, Y., Ji, Q., & Zhang, D. (2019). Impacts of China-US trade conflicts on the energy sector. *China Economic Review*, *58*, 101360. https://doi.org/10.1016/j.chieco.2019.101360

Yan, Z., Xiao, Y., Cheng, L., Chen, S., Zhou, X., Ruan, X., Li, M., He, R., & Ran, B. (2020). Analysis of global marine oil trade based on Automatic Identification System (AIS) data. *Journal of Transport Geography*, *83*, 102637. https://doi.org/10.1016/j.jtrangeo.2020.102637

Yang, D., Wu, L., Wang, S., Jia, H., & Li, K. X. (2019). How big data enriches maritime research – a critical review of Automatic Identification System (AIS) data applications. *Transport Reviews*, *39*(6), 755–773. https://doi.org/10.1080/01441647.2019.1649315

Yang, Y. (2018). Transport infrastructure, city productivity growth and sectoral reallocation: Evidence from China. *IMF Working Paper No. 18/276*. https://www.imf.org/en/Publications/WP/Issues/2018/12/11/Transport-Infrastructure-City-Productivity-Growth-and-Sectoral-Reallocation-Evidence-from-46456

Zhang, C., Bin, J., Wang, W., Peng, X., Wang, R., Halldearn, R., & Liu, Z. (2020). AIS data driven general vessel destination prediction: A random forest based approach. *Transportation Research Part C: Emerging Technologies*, *118*, 102729. https://doi.org/10.1016/j.trc.2020.102729

Zhang, W., Goerlandt, F., Montewka, J., & Kujala, P. (2015). A method for detecting possible near miss ship collisions from AIS data. *Ocean Engineering*, *107*, 60–69. https://doi.org/10.1016/j.oceaneng.2015.07.046