Title
Three Papers in Labor Market and Commodity Market

Author
Zhang, Yunxiao

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

THREE PAPERS IN LABOR MARKET AND COMMODITY MARKET

A dissertation submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ECONOMICS

by

Yunxiao Zhang

September 2021

The Dissertation of Yunxiao Zhang
is approved:

Professor Grace Gu, Chair

Professor Kenneth Kletzer

Professor Robert Fairlie

Professor Brenda Samaniego de la Parra

Peter Biehl
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

THREE PAPERS IN LABOR MARKET AND COMMODITY MARKET

by

Yunxiao Zhang

This dissertation presents three empirical studies on topics in labor market and global commodity prices. The first chapter analyzes how consumption, spending and labor market decisions vary across the life cycle, and their implications for the aggregate economy. The second chapter discovers the dynamic adjustments of 49 world commodity prices in response to innovations in the nominal and real shocks, and the third chapter reveals how commodity prices are affected by information from the news in high-frequency trading.

The first chapter studies the effect of aging on service consumption quantity and price in the United States. In particular, I ask, do older households consume more service goods? Can aging explain the rise of service price? I investigate the impact of age profile on service consumption and service price in the U.S. using household survey data and metropolitan statistical areas (MSAs) level variations, respectively. The results show that after controlling for household income, as people age, they consume progressively more health insurance and medical services and less food away from home. Furthermore, there is a positive correlation between service price and old-age dependency ratio1 in the long run across MSAs. I then build a two-period two-sector OLG model with an education entry cost to explain these results. In this model, age affects service price through two channels: 1) on the demand side, labor markets with older populations tend to demand more services; 2) on the supply side, the education friction would cause worker mobility from the service sector to the manufacturing sector, which further increases the service price.

The second chapter is a group project with Professor Hyeongwoo Kim. We study dynamic adjustments of 49 world commodity prices in response to innovations in the nominal exchange rate and the world real GDP. After we estimate the dynamic elasticity of the prices with respect to these shocks, we obtain the kernel density of our estimates to establish stylized facts on the adjustment process of the commodity price toward a new equilibrium path. Our empirical findings imply, on average, that the law of one price holds in the long-run, whereas the substantial degree of short-run price rigidity was observed in response to the nominal exchange rate shock. The real GDP shock tends to generate substantial price fluctuations in the short-run because adjustments of the supply can be limited, but have much weaker effects in the long-run as the supply eventually counterbalances the increase in the demand. Overall, we report persistent long-lasting effects of the nominal exchange rate shock on commodity prices relative to those of the real GDP shock.

The third chapter is a group project with Yifei Sheng and Zijing Zhu. To uncover the news impact on the price of WTI crude oil futures, the third chapter applies supervised and unsupervised machine learning algorithms to conduct news sentiment and topic analysis. With the assumption that the crude oil futures market is efficient enough to respond quickly to new information, this chapter obtains high-frequency price and news from the Bloomberg terminal. Using results from logistic regression and K-means clustering, this chapter defines the positive score and topic for each news article as inputs for the final logistic regression. The regression results show that the "World Crude Oil" news is more positively correlated with price increase than other topics. Moreover, the "WTI Crude Oil" news has the highest correlation with the price increase as the positive score increases.

To myself and my family,

practice makes perfect.

Acknowledgments

At this part, I want to express my sincere appreciation to the people who helped me, showed their kindness, and offered me endless support during the pursuit of my Ph.D. journey.

First, I want to thank my advisor, Professor Grace Gu, for her excellent guidance and unbelievable patience throughout the years. Her patient and optimistic personality inspired me to overcome difficulties and stay positive when I have troubles during my research. As a talented and young economist, I learned a lot from her. It is my great pleasure to study with Grace in these years.

I want to give special thanks to my committee members, Professor Kenneth Kletzer, Professor Brenda Samaniego de la Parra, and Professor Robert Fairlie for their efforts and tremendous help in my research. I appreciate all their contributions for the time, effort, and ideas that made my research experience productive and unforgettable in my study.

I also want to thank all other professors who taught me the coursework with valuable knowledge and give me great feedback on my research. Furthermore, I would like to thank all the participants in the macro workshops and the job market practice talks for their helpful comments and suggestions.

I want to thank the program coordinator Sandra, who has always been there to help with anything. I want to thank my friendly and supportive cohorts. I cannot make this come true without their company. We have been together to the same goal for years, and it is such a great honor for me to spend the years with them.

Last but not least, I would like to thank my family for all their love, kindness, and encouragement both physically and mentally. A special thank you to my beloved, encouraging partner NZ, who has given me tremendous support and help during the years. Doing a Ph.D. is not an easy thing, but I am glad that I overcome all the difficulty ad make the dream come

true. I would keep all the good memories and continue on another new journey.

Chapter 1

The Effect of Aging on Service Consumption and Price

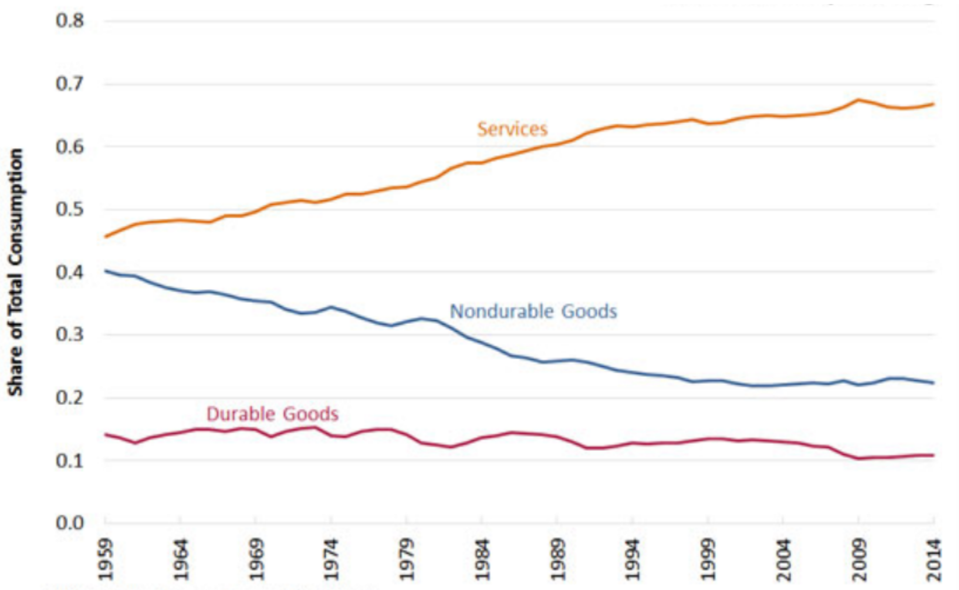
1.1 Introduction

As shown in Figure 1.1, from 1959 to 2014 in the United States, the share of services in the total expenditure increased from 46% to 67%, while the share of nondurable goods decreased from 40% to 22% during the same period. In other words, U.S. consumers have spent more on services than durable and nondurable goods over time.

However, since we know that total consumer expenditure is equal to the product's price multiplied by the quantity, the increase in service expenditure is not necessarily due to growth of quantity, or in other words, real consumption share. Consumers might have spent more on services simply due to higher prices for services even if the quantity consumed has remained constant.

Figure 1.2 shows real consumption share for each category, with the consumer expenditure in each category adjusted by its associated inflation index. Figure 1.2 shows a different

Figure 1.1: U.S. consumption spending share, 1959-2014



pattern compared to Figure 1.1. Based on Figure 1.2, the real consumption share of services in the United States has remained relatively stable between 62% and 70% from 1959 to 2014. Figure 1.2 also indicates that in terms of the relative share of real consumption, U.S. consumers have tended to consume more durable goods and fewer nondurable goods over time.

Over the same period, the U.S. population has aged. Figure 1.3 captures this trend. The number of Americans aged 65 and over was 16.2 million in 1960 and increased to 49.2 million in 2016. People aged 65 and over represented 8.9% of the population in the year 1960 but increased to be 15.2% of the population in 2016.[1] Can population aging influence the price and quantity demand for services? On the one hand, the life-cycle hypothesis suggested that consumer spending patterns vary across different age groups (e.g., Chen and Chu (1982) and Harris and Blisard (2002)). On the other hand, societies with different age profiles might have different price patterns. However, to the best of my knowledge, existing literature has focused

[1]Statistics are from "A profile of older Americans: 2017" (U.S. Department of Health and Human Services, Administration on Aging, 2017).

Figure 1.2: U.S. real consumption spending share, 1959-2014

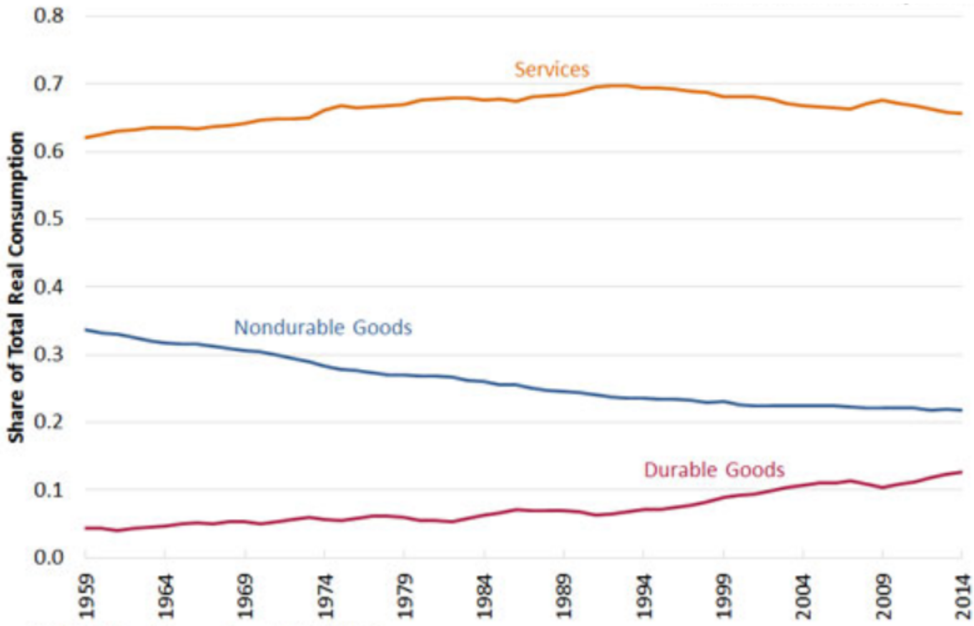
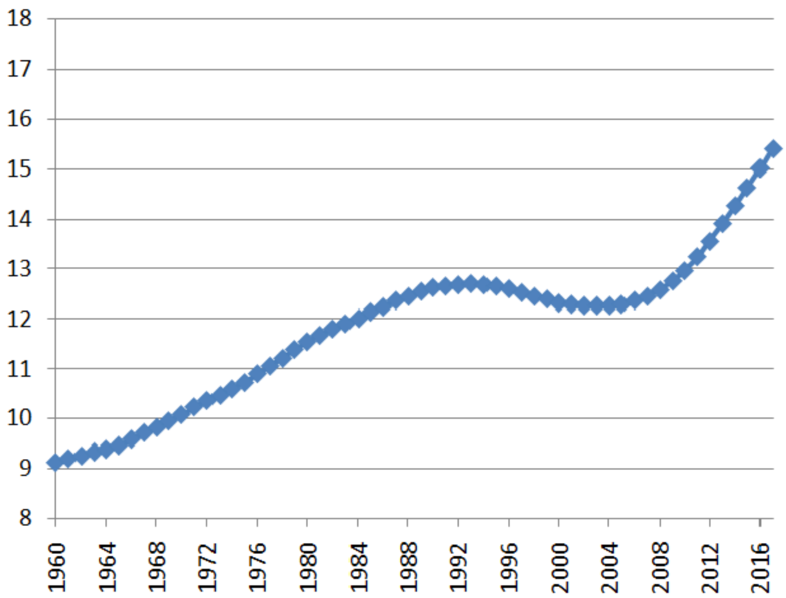


Figure 1.3: Percent of persons aged 65 and over in the United States



on the service quantity or service price, but there is little evidence analyzing the potential interactions of both service quantity and price. It is important to investigate the effect of population aging on service consumption and price simultaneously so that we can have a better understanding of the general equilibrium in order to look at both the supply and demand sides.

In this paper, I investigate the impact of age profile on service consumption and service price in the United States using household survey data and metropolitan statistical areas (MSAs) variation in price, respectively. In particular, I ask, do older households consume more service goods? Can aging explain the increasing trend in service price?

This paper contributes to the previous literature. First, previous studies focused on the effect of aging on goods consumption and the effect of aging on relative service prices separately. However, in this paper, I consider the impact of age profile on service consumption as well as the service price within the United States simultaneously. It is significant to combine the study of age profile on price and quantity in order to determine whether the observed changes are due to aging's impact on supply or demand. Second, in order to better understand the empirical results, I build a two-period two-sector OLG model with entry costs.

In this paper, first, I use the U.S. household survey data from the Consumer Expenditure Survey (CEX) from the years 2006-2014 to investigate the effects of aging on service consumption. The results show that after controlling for household income and other characteristics, as people age, they consume progressively more health insurance and medical services and less food away from home.

Second, I investigate the impacts of age profile on service price across the U.S. MSAs using household survey data from the American Community Survey (ACS) for the period 2008-2016. The results show that there is a positive correlation between service price and the old-age dependency ratio[2].

[2] Old-age dependency ratio is defined as the ratio of the population aged 65 and over to the working-age

To further investigate the impact of population aging on the rise of equilibrium service consumption and service price, I build a two-period two-sector OLG framework to explain the empirical results. The channel I implement is education entry costs, which assume that younger generations are required to attend school for work training. Young workers in both manufacturing and service sectors need training, but training to work in the service sector has a higher cost. This education entry cost would result in fewer workers in the service sector and further increase the service price. The two-period OLG model captures the impact of the demographic structure. A representative agent maximizes her lifetime utility. Young and old have different preferences for services. In the first period, the young pay the entry cost and work. The young also consume and save. In the second period, the old consume the savings and do not work. Through comparative statics, this two-period two-sector model allows me to separately analyze the contribution of aging to a demand (preferences) and a supply (education training) channel.

The remainder of this paper is organized as follows: section 1.2 presents the related literature. Section 1.3 shows the study of aging on service consumption, including the empirical methodology, data, and preliminary results. Section 1.4 shows the study of aging on service price, including the empirical methodology, data, and preliminary results. Section 1.5 shows the framework of the model. Section 1.6 concludes and talks about future steps.

1.2 Literature Review

This paper mainly relates to three streams of research: (i) studies on population aging with goods consumption; (ii) studies on age profile with the relative price of service (non-tradable goods); (iii) studies on population aging with the growth of the service sector.

A number of previous literature studied household consumption patterns of older Amer-

---

population aged 15 to 64.

icans. Researchers have investigated consumption patterns in terms of both total dollar expenditures and expenditure shares of different consumption categories. Some studies on how the elderly allocate their expenditures show that older households spend more of their income on basic needs than do younger households. Compared with the younger households, the elderly spend more on housing, food, and health care, and less on clothing, transportation, and household furnishings (Yung-Ping Chen and Kwang-Wen Chu, 1982; Gong-Soog Hong and Soo Yeon Kim, 2000; J. Michael Harris and Noel Blisard, 2002 and many others). Previous papers mainly focus on the impact of aging on different consumption categories, which can be considered as the effect of aging on service quantity after controlling the inflation index. While this paper investigates the effect of aging on service consumption equilibrium in quantity and price.

Since service goods can be seen as non-tradeable goods, most studies about the service price are mainly about the relative price of non-tradeable goods relative tradeable goods. The literature distinguishes three structural determinants of the relative price of nontradeables (Bergstrand, 1991): productivity differentials (Balassa-Samuelson), relative factor endowments (Bhagwati), and relative demand shocks (Bergstrand). Important determinants of the relative price beyond above in the literature are government demand (De Gregorio et al. 1994, Galstyan and Lane 2009), net foreign assets (Lane and Milesi-Ferretti 2004, and Christopoulos et al., 2012) and imperfect competition (Coto-Martinez and Reboredo, 2012). Bettendorf and Dewachter (2007) focus on the effect of changes in the population age structure on the relative price of non-tradable goods for 16 OECD countries from 1970 to 2002 using both supply and demand factors. However, their empirical findings remain inconclusive and insignificant in the majority of cases. In recent literature, Groneck and Kaufmann (2017) present evidence that elderly people consume more non-traded services relative to people of working-age using a sample of 15 OECD countries between 1970 and 2009. They find that this demand shift increases the relative price of non-tradable goods and thereby causes real exchange rates to appreciate. So far, most studies about the effect of aging on service goods focus on a variety of countries. It would be significant if using U.S. household survey data to demonstrate this point. In this paper, I apply the context from this multi-country case to investigate the effect of aging on service prices within the U.S.

This paper also relates to the literature of population aging with the growth of the service sector. Previous papers of investigating the demographic profile of consumption expenditures mainly use frictions on factor markets, such as imperfect intersectoral labor mobility, as proposed by Horvath (2000) and Cardi and Restout (2013). While this paper tries to apply educational friction as entry costs to investigate the effect of aging on the rise of the equilibrium service price.

1.3 Service Consumption

1.3.1 Data

The household-level data are drawn from the Bureau of Labor Statistics (BLS) Consumer Expenditure Survey (CEX), which provides detailed data on consumption expenditure, income, socioeconomic and demographic characteristics of U.S. households since the early 1980s. The CEX survey is a rotating panel. Each household is interviewed every three months over four consecutive quarters to obtain a year's worth of data and every quarter 20% of the sample is replaced by new households. In total, around 5000 households participate in the survey each quarter.

In this study, I apply the 2006-2014[3] data from the CEX to investigate the influence of aging on service consumption using a cross-sectional empirical specification. For each household, I define the measurement of consumption in each service goods ($c_{j,t}$) as the annual consumption

[3] Consumer price index (p_t) for some services such as vehicle insurance is limited, therefore the data is restricted to 2006-2014

summed up the consumption across all four interviews, dropping households with fewer than the four complete interviews. For each household, there are 7 consumption categories of service goods, including food away from home, education, vehicle insurance, vehicle maintenance and repairs, public transportation, health insurance, medical services. The income measure is the income of households after taxes for the previous 12 months, taken from the CEX family file corresponding to the final interview of a household, so it covers the same period as the consumption measure.

The Consumer Price Index (p_t) of each service good is taken from the BLS website. This Consumer Price Index (p_t) is not seasonally adjusted and is the component for urban consumers. The dollar figures are adjusted to 1982-84 dollars.

To give an impression of the magnitude and evolution over time of the data, I present some summary statistics in Table 1.1. I show means, observations, and standard deviations at the beginning and end of the observation period. From Table 1.1, we can see that the average consumption for each household in the category of food away from home, education, vehicle maintenance and repairs, public transportation, health insurance, and medical service increase in 2014 compared to 2006. While the average consumption for each household in vehicle insurance does not change much in the year 2014 relative to the year 2006.

1.3.2 Empirical Methodology

Following Erlandsen and Nymoen (2008) and Ma (2017), in order to investigate the impact of aging on service consumption, I estimate each service consumption category in the following form:

$$log(c_{j,t}/p_t) = \beta_0 + \beta_1 log(income_{j,t}) + \delta_k age_{k,t} + \beta_2 control_{j,t} + \epsilon_{j,t} \tag{1.1}$$

Table 1.1: summary statistics on service consumption

Variable	2006			2014		
	Mean	Std. Dev.	Obs	Mean	Std. Dev.	Obs
age_{jt}	51.14	16.31	2501	53.07	16.73	2185
$income_{jt}$	54489.54	69097.39	2501	61190.71	72901.35	2185
$c_{j,t}$						
Food away from home	1715.17	2211.75	2501	2541.50	2725.73	2185
Education	875.67	3697.11	2501	1077.40	6265.65	2185
Vehicle insurance	975.78	1050.39	2501	962.19	1059.85	2185
Vehicle maintenance	722.80	1063.29	2501	761.67	1267.62	2185
Public transportation	81.61	389.48	2501	103.52	408.50	2185
Health insurance	1578.19	1893.23	2501	3138.96	3755.02	2185
Medical service	733.70	1603.73	2501	862.47	1870.77	2185

Note: $income_{j,t}$ and each service consumption $c_{i,j,t}$. are measured in U.S. dollars. The dollar figures are adjusted to 1982-84 dollars.

Where j is household. $k = 1, 2, ..., 5$ defines each age group. For each household, there are 7 consumption categories of service goods, including food away from home, education, vehicle insurance, vehicle maintenance and repairs, public transportation, health insurance, and medical services. For age groups, the five dummy variables are age 25-34, age 35-44, age 45-54, age 55-64, age 65 and over, with the age group under 25 the omitted classification.

In Equation 1.1, the dependent variable is the quantity consumed in services of each category for a household. I calculate consumption in quantities by deflating each household's expenditures on each consumption good by the corresponding Consumer Price Index (p_t). The explanatory variables in the estimation process include 5 age dummy variables and the annual total nominal income of the households. Age is defined as the age of reference person[4]. $log(income_{j,t})$ is the log of total income after tax for the past 12 months within a household. Control variables include the number of kids under 18, race categories, sex and education cate-

[4]The reference person of the consumer unit is the first member mentioned by the respondent when asked to "Start with the name of the person or one of the persons who owns or rents the home".

gories of the reference person, homeownership status as well as the geographical states.

1.3.3 Preliminary Empirical Findings

I first estimate Equation 1.1 for each consumption category using a pooled regression across years, then obtain the total consumption expenditure as the sum of seven individual consumption expenditures.

Table 1.2 shows the pooled results across years of the regression for Equation 1.1. Different service goods show different consumption patterns across the age groups. Column I shows that expenditure on food away from home is negatively and significantly correlated with age profile, indicating that as people get older, they spend less on food away from home, which confirmed the results of previous studies (J. Michael Harris and Noel Blisard, 2002). Expenditure on health insurance (column VI) and medical service (column VII) is positively and significantly correlated with age profile. As people age, they consume progressively more quantity on health insurance and medical services. While, for services such as education (column II), vehicle insurance (column III), vehicle maintenance and repairs (column IV), there is no clear increasing or decreasing trend between consumption quantity and age profile. The correlation between each service consumption with income level is positive, which means a rise in income leads to an increase in service consumption quantity, which is not surprising.

Table 1.3 shows the pooled regression results across years for total service consumption. I define the total service consumption as the sum of all 7 consumption categories. I then deflate each category with the corresponding Consumer Price Index (p_t) and sum all of them. Results show that there is a positive correlation between consumption in services quantity and households whose age is 65 and over, indicating that people who are 65 and over tend to consume more quantity on service goods.

Table 1.2: Baseline pooled regression results for each service consumption

VARIABLES	(I) food	(II) edu	(III) vehins	(IV) vehmain	(V) transport	(VI) insurance	(VII) medical
age: 25-34	-0.311***	-1.139***	0.454***	0.088	-0.192*	0.686***	0.285**
	(0.077)	(0.167)	(0.147)	(0.121)	(0.110)	(0.159)	(0.131)
age: 35-44	-0.400***	-0.795***	0.354**	0.010	-0.152	1.033***	0.495***
	(0.078)	(0.165)	(0.147)	(0.121)	(0.109)	(0.159)	(0.130)
age: 45-54	-0.476***	-0.059	0.508***	0.298**	-0.041	1.284***	0.824***
	(0.077)	(0.164)	(0.144)	(0.119)	(0.108)	(0.156)	(0.129)
age: 55-64	-0.723***	-1.134***	0.380***	0.253**	-0.224**	1.736***	1.151***
	(0.080)	(0.164)	(0.147)	(0.120)	(0.108)	(0.158)	(0.131)
age: 65 and over	-1.093***	-1.831***	0.032	-0.356***	-0.438***	3.726***	1.054***
	(0.082)	(0.160)	(0.146)	(0.121)	(0.106)	(0.151)	(0.130)
income	0.394***	0.238***	0.412***	0.391***	0.038***	0.393***	0.425***
	(0.016)	(0.021)	(0.021)	(0.018)	(0.014)	(0.022)	(0.020)
Constant	-2.947***	-6.356***	-7.473***	-6.547***	-4.458***	-6.469***	-9.243***
	(0.453)	(0.484)	(0.537)	(0.439)	(0.431)	(0.603)	(0.508)
Control	yes	yes	yes	yes	yes	yes	yes
Observations	17,101	17,101	17,101	17,101	17,101	17,101	17,101
R-squared	0.240	0.200	0.160	0.215	0.161	0.229	0.208

Note: Column(I)-Colume(VII) are the results of regressions for each service industry, which is food away from home, education, vehicle insurance, vehicle maintenance and repairs, public transportation, health insurance and medical services, respectively. Baseline reference group for age is under 25. Asterisks mark significance at 10% (*), 5% (**), 1% (***). The standard error is robust standard error.

Table 1.3: Baseline pooled regression results for aggregate service consumption

VARIABLES	services
age: 25-34	-0.144***
	(0.043)
age: 35-44	-0.131***
	(0.043)
age: 45-54	0.019
	(0.043)
age: 55-64	-0.033
	(0.045)
age: 65 and over	0.113***
	(0.043)
income	0.253***
	(0.009)
Constant	-0.851***
	(0.305)
Observations	17,101
R-squared	0.346

Note: Baseline reference group for age is under 25. Asterisks mark significance at 10% (*), 5% (**), 1% (***). The standard error is robust standard error.

1.3.4 Robustness Tests

In this section, I present the results of robustness tests using a different sample to confirm the results I get.

Besides The Consumer Expenditure Survey, another national survey that has collected data on the consumption expenditures over a long period is the Panel Study of Income Dynamics (PSID). Historically, this survey collected information only on food and housing expenditures. Beginning in 1999, the PSID had the questions about other expenditures, including spending on transportation, health care, education, utilities, and childcare. With this expanded set of questions on consumption expenditures, the PSID covered more than 70 percent of the total outlays measured in the Consumer Expenditure Survey. This makes PSID a perfect sample substitute to test the results I have.

Table 1.4 presents the pooled regression results across years for total service consumption. I define the total service consumption as the sum of all 7 consumption categories same as in the previous study. It include food away from home, education, vehicle insurance, vehicle maintenance and repairs, public transportation, health insurance, and medical services. I then deflate each category with the corresponding Consumer Price Index (p_t) and sum all of them. Same to use the Consumer Expenditure Survey, results show that there is a positive correlation between consumption in services quantity and households whose age is 65 and over, indicating that people who are 65 and over tend to consume more quantity on service goods.

1.4 Service Price

1.4.1 Data

To investigate the effect of aging on service prices across different cities, the challenge is to find the best measurement of the price index. We can not choose the regional CPI data provided by the BLS in this scenario, since these data only measure how prices change over time and not the relative prices in regions at any point in time. One possible data resource is ACCRA (American Chamber of Commerce Researchers Association) index of U.S. urban prices, which is widely used in city comparisons by researchers. However, this dataset is only about the goods sector. Koo, Phillips, and Sigalla (2000) argued that the potential weaknesses of the ACCRA indexes include sampling error, sampling bias, aggregation bias, and substitution bias.

In this case, I use Regional Price Parities (RPPs) as the price index, which is released from the Bureau of Economic Analysis (BEA). RPPs are price indexes that measure geographic price level differences such as states or metropolitan statistical areas (MSA) for one period in time within the United States. For example, the New York City metropolitan area had a 2014 RPP of 122.3, which means that NYC is about 22.3% more expensive than the national average.

Aten (2005) and Aten and Martin (2012) carefully explain the calculation of RPPs. The calculation of RPP is based on RPP expenditure weights, which is derived from Consumer Expenditure (CEX) Survey data from the Bureau of Labor Statistics (BLS). These results are adjusted to incorporate rents expenditures from the American Community Survey (ACS) of the Bureau of Census. Finally, the weights are balanced to reflect the commodity distribution of Personal Consumption Expenditures (PCE) of the Bureau of Economic Analysis.

To investigate the age profile, average household income, and unemployment rate within the metropolitan statistical area (MSA), I use the data from American Community Survey (ACS) of the Bureau of Census. I then match the data of RPPs and other data from the

American Community Survey (ACS) together if they have the same zip code. 98 metropolitan statistical areas (MSA) are included in the year 2008 and 107 metropolitan statistical areas (MSA) are included in the year 2016, which are all available cities under the classification of metropolitan statistical areas (MSA) in American Community Survey (ACS) from the year 2008 to 2016.[5] I then investigate the effect of aging on service price using the combined data set with the cross-sectional empirical specification.

Table 1.5 presents the summary statistics of all the variables applied to investigate the effect of aging on service price. I show means, observations, and standard deviations at the beginning and end of the observation period as well as all periods to have an overview of the data. From Table 1.5, we can see that the average Regional Price Parities (RPPs), the average median age, and the average income for each MSA do not change much in the year 2008 relative to the year 2016. While the average old-age dependency ratio increases by 4.6% from the year 2008 to the year 2016.

1.4.2 Empirical Methodology

To capture the relation of population aging on the equilibrium of both service consumption and price, I develop an empirical specification to investigate the effect of aging on service price across the U.S. metropolitan statistical areas (MSA) using household survey data from American Community Survey (ACS) for the year 2008 to 2016. Following Bettendorf and Dewachter (2007), the cross-sectional equation to be estimated for the service price with state fixed effects γ_j in MSA i is:

$$log(rpp_i) = \beta_0 + \beta_1 oldratio_i + \beta_2 log(income_i) + \beta_3 log(employment_i) + \gamma_j + \epsilon_i \tag{1.2}$$

The dependent variable is the log of average Regional Price Parities (RPPs) across

[5] Regional Price Parities (RPPs) only publicly available from 2008-2016.

Table 1.4: Baseline pooled regression results for total service consumption from PSID 2007-2013

VARIABLES	services
age: 25-34	-0.126***
	(0.023)
age: 35-44	-0.114***
	(0.023)
age: 45-54	0.016
	(0.024)
age: 55-64	-0.087***
	(0.025)
age: 65 and over	blue 0.117***
	(0.026)
income	0.479***
	(0.008)
Constant	-2.001***
	(0.089)
Observations	34,124
R-squared	0.318

Baseline reference group for age is under 25. The standard error is robust standard error.

Table 1.5: Summary statistics on service price

Variable	2008			2016			Pooled		
	Mean	Std. Dev.	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.	Obs
$rpp_{i,t}$	100.96	26.37	98	99.40	28.67	107	99.45	26.59	921
$oldratio_{i,t}$ (%)	18.78	4.69	98	23.36	6.13	107	20.86	5.77	921
$age\_median_{i,t}$	36.56	3.28	98	37.83	3.65	107	37.23	3.62	921
$income_{i,t}$ (in 2016 \$)	60963.01	10822.9	98	60032.95	11819.87	107	57690.13	10653.43	921
$employment_{i,t}$ (million)	92.85	118.09	98	99.99	133.05	107	93.39	121.39	921

years for service goods from the Bureau of Economic Analysis (BEA). On the right-hand side, *oldratio* measures the old-age dependency ratio, which is defined as the ratio of the population 65 and over to the working-age population between ages 15 and 64. $log(income)$ is the log of average household income in the past 12 months. *employment* is the number of employed in each MSA, which captures the supply side of the determinant of the service price.

As we may see from Table 1.5, the time trend for the variable in the left-hand side and right-hand side in Equation 1.2 is sort of different. To avoid the influence of different trend patterns on the effect of the results, I then run a cross-sectional regression with averages in Equation 1.2.

1.4.3 Preliminary Empirical Findings

Table 1.6 reports the baseline regression results for Equation 1.2 with state fixed effects. The results suggest a positive correlation between service price and the old-age dependency ratio. Table 1.6 also suggests a positive correlation between service price with the log of the number employed and the log of average household income, respectively. These two results are not surprising, since they capture the basic demand and supply factor of influencing service price.

1.4.4 Other results

In section 1.4.3, I have shown that the price of services varies across MSAs and that this variation is correlated with heterogeneity in old age dependence ratios. Next, I want to check whether this phenomenon is exclusive to the service sector or if it also affects the goods sector. In order to do so, the estimated cross-sectional equation for the relative service price

Table 1.6: Baseline regression results for service price

VARIABLES	state fixed effects
oldratio	0.003*
	(0.002)
logincome	0.941***
	(0.079)
logemployment	0.059***
	(0.020)
Constant	-7.055***
	(0.750)
Observations	107
R-squared	0.946

Note: Asterisks mark significance at 10% (*), 5% (**), 1% (***). The standard error is cluster std error by state.

with state fixed effects γ_j in MSA i is as followings:

$$log(rpp_{service;i})log(rpp_{goods;i}) = \beta_0 + \beta_1 oldratio_i + \beta_2 log(income_i) + \beta_3 logemployment_i + \gamma_j + \epsilon_i \tag{1.3}$$

Table 1.7 reports the regression results for Equation 1.3 both with state fixed effects and without state fixed effects.The results suggest a positive correlation between relative service price and old-age dependency ratio when we do no consider state fixed effects. Since state fixed effects will capture more variations across the MSA, this result is pretty similar to the baseline results we get in Table 1.6.

As we know, the price adjustments may take time. To confirm if the result is the same in the short period, I take the lag variables of the old-age dependence ratio to test the result:

$$\triangle log(rpp_{i,t}) = \beta_0 + \beta_1 \triangle oldratio_{i,t-1} + \beta_2 \triangle log(income_{i,t}) + \beta_3 \triangle logemployment_{i,t} + \gamma_i + \delta_t \tag{1.4}$$

Table 1.8 reports the regression results for Equation 1.4 using the lag of old-age dependency ratio with state fixed effects. When we take into consider the first and second lag of old-age

Table 1.7: Regression results for relative service price

VARIABLES	(1) state fixed effects	(2) no state fixed effects
oldratio	0.001	0.002***
	(0.000)	(0.001)
logincome	0.188***	0.222***
	(0.017)	(0.024)
mlogemployment	0.010**	0.011***
	(0.004)	(0.004)
Constant	-1.294***	-1.684***
	(0.123)	(0.205)
Observations	107	107
R-squared	0.938	0.690

Note: Asterisks mark significance at 10% (*), 5% (**), 1% (***). Column (1) is for regression with state fixed effects with cluster std error by state; column (2) is for regression without fixed effects with cluster std error by state.

Table 1.8: Regression results for service price with lag old-ratio

VARIABLES	(1) with state fixed effects	(2) with state fixed effects
oldratiolag1	-0.001	-0.001
	(0.003)	(0.003)
g_logincome	0.063***	0.063***
	(0.017)	(0.019)
g_logemployment	0.045**	0.034*
	(0.021)	(0.020)
oldratiolag2		0.000
		(0.003)
Constant	-0.005*	-0.008***
	(0.003)	(0.002)
Observations	707	601
R-squared	0.414	0.241

Note: (1) and (2) columns are for state fixed effects with cluster std error by state for $\triangle oldratio_{i,t-1}$ and $\triangle oldratio_{i,t-2}$, respectively dependency ratio, the results become not significant. It may suggest that the price rigidity can have an influence for the price adjustment in the short run.

1.5 Model

The empirical part provides evidence for the positive correlation between the rise of the service price and old-age dependence ratio in the U.S. The regression results also show a positive correlation between service consumption and old households (age 65 and over) compared to the young. In this section, I build up a model to suggest a potential mechanism behind these correlations.

The model is an OLG model. In this simple set-up, the model has two generations: young and old. There are two sectors in the economy: the manufacturing and the service sectors. These two sectors produce two consumption goods: manufacturing and service goods. There is no capital in this model, labor is the only input. Individuals only work when they are young. In this model, I assume: (i) workers need the training to work in both sectors but the cost of this education is higher in the service sector; (ii) old and young have different preferences for services; (iii) old people continue to increase in the economy. There are two channels through which age affects service price in this model: 1) in the demand side, labor markets with older populations tend to demand more services; 2) in the supply side, the education friction would cause worker mobility from service sector to the manufacturing sector, then further increase the service price.

According to the model, a young person works, goes to school to get the education for work training, and earns wages. When she retires, she simply consumes savings. I assume that the representative household is endowed with 1 unit of labor when young, and the household can allocate this unit of labor to work in both manufacturing sector for a fraction and service sector for the remaining fraction.

To better understand the mechanism, I begin with a simple version with educational friction, which means the young generation has to pay education cost in school, in the meanwhile, she can work.

Household

$$max[logC_{y,t} + \beta logC_{o,t+1}] \tag{1.5}$$

where $C_{i,t}$ $i \in \{y, o\}$ is the composite (CES) of consumption in manufacturing and service goods:

$$C_{i,t} = [\omega_{i,s}{}^{\frac{1}{\varepsilon}} c_{i,st}^{\frac{\varepsilon-1}{\varepsilon}} + \omega_{i,m}{}^{\frac{1}{\varepsilon}} c_{i,mt}^{\frac{\varepsilon-1}{\varepsilon}}]^{\frac{\varepsilon}{\varepsilon-1}} \tag{1.6}$$

where $c_{i,st}$ and $c_{i,mt}$ the consumption in services and in manufacturing goods; $\omega_{i,s}$, $\omega_{i,m}$ the relative weights of service goods and manufacturing goods consumption. I assume $\omega_{o,s} > \omega_{o,m}$, which is based on my empirical results that old households (age 65 and over) tend to demand more services.

Budget constraints are:

$$c_{y,mt} + p_{st}c_{y,st} + a_{y,t} = (w_{st} - e_s)S_{st}N_{yt} + (w_{mt} - e_m)S_{mt}N_{yt} + \Pi_s + \Pi_m \tag{1.7}$$

where, $S_{st} + S_{mt} = 1$; Π_s and Π_m is the firm's profit for service and manufacturing sector, respectively.

$$c_{o,mt+1} + p_{st+1}c_{o,st+1} = (1 + r)a_{y,t} \tag{1.8}$$

where p_{st} is the relative price of service goods; w_{st} and w_{mt} are wages for service and manufacturing workers, respectively; e_s and e_m are education cost for entering service and manufacturing industry, respectively; I also assume e_s ¿ e_m; S_{st} and S_{mt} are labor share working in service and manufacturing industry, respectively; N_{yt} is the total young generation.

The household choose consumption and labor share to maximize their utility taking

Π_s and Π_m as given, we have:

$$(\omega_{i,s}c_{i,mt}\omega_{i,m}c_{i,st})^{\frac{1}{\varepsilon}} = p_{st}, i \in \{y, o\} \tag{1.9}$$

$$w_{st} - e_s = w_{mt} - e_m \tag{1.10}$$

Where, Equation 1.9 interprets the relationship between demand and relative price p_{st}.

Firm Households own the firm. In each sector, firms use labor to produce manufacturing/service goods. The production functions are:

$$Y_{jt} = A_j(L_{jt})^{\alpha}, j \in \{s, m\} \tag{1.11}$$

where, $L_{jt} = S_{jt}N_{yt}, j \in \{s, m\}$

Firms maximize their profits, so we have

$$max\Pi_s = p_{st}A_s(L_{st})^{\alpha} - w_{st}L_{st} \tag{1.12}$$

$$max\Pi_m = A_m(L_{mt})^{\alpha} - w_{mt}L_{mt} \tag{1.13}$$

By taking F.O.Cs of each sector, we have:

$$w_{st} = \alpha p_{st}A_s{L_{st}}^{\alpha-1} \tag{1.14}$$

$$w_{mt} = \alpha A_m{L_{mt}}^{\alpha-1} \tag{1.15}$$

Equilibrium Taking $L_{jt} = S_{jt}N_{yt}, j \in \{s, m\}$ into Equation 1.14 and 1.15, we have:

$$w_{st} = \alpha p_{st}A_s(S_{st}N_{yt})^{\alpha-1} \tag{1.16}$$

$$w_{mt} = \alpha A_m(S_{mt}N_{yt})^{\alpha-1} \tag{1.17}$$

Taking the condition of wage equality $w_{st} - e_s = w_{mt} - e_m$ into Equation 1.16 and 1.17, we have:

$$\alpha N_{yt}^{\alpha-1}(p_{st}A_s{S_{st}}^{\alpha-1} - A_m{S_{mt}}^{\alpha-1}) = e_s - e_m \tag{1.18}$$

The resource constrains are:

$$Y_{st} = c_{y,st} + c_{o,st} \tag{1.19}$$

$$Y_{mt} = c_{y,mt} + c_{o,mt} \tag{1.20}$$

For the population, I assume N_{ot} is the old total generation and $N_{ot} + N_{yt} = 1$.

Comparative statics From Equation 1.18, we could see that there is a relation between N_{yt} and p_{st}, then I could change the relative entry costs and seeing how this affects the share of workers in each sector and the relative price. Using the comparative statics, we can have:

$$d_{p_{st}} d_{e_s - e_m} = 1\alpha N_{yt}^{\alpha-1} A_s {S_{st}}^{\alpha-1} > 0 \tag{1.21}$$

Equation 1.21 suggests that when the gap of cost $e_s - e_m$ increases, p_{st} will also increase; when the gap of $e_s - e_m$ decreases, p_{st} will also decrease.

$$d_{p_{st}} d_{N_{yt}} = (\alpha - 1)(e_s - e_m) N_{yt}^{\alpha-2} \alpha N_{yt}^{\alpha-1} A_s {S_{st}}^{\alpha-1} < 0 \tag{1.22}$$

Equation 1.22 suggests that when the size of young households N_{yt} increases, p_{st} will decrease; when size of young households N_{yt} decreases, p_{st} increase. These analytical results confirm the results from price analysis, which means when the total population is stable, the more old generations tend to have a positive correlation between relative service price.

1.6 Conclusions

This paper studies the effect of aging on service consumption quantity and price in the United States. In particular, I ask, do older households consume more service goods? Can aging explain the rise of service price? To answer these questions, I investigate the impact of age profile on service consumption and service price in the U.S. using household survey data and metropolitan statistical areas (MSAs) level variations, respectively.

The results show some interesting features over the effect of aging on service consumption. Aligned with some previous empirical literature, in this study, I found that after controlling for household income and other characters, as people ages, they consume progressively more quantity on health insurance and medical services and less on food away from home. For the service price, the empirical results suggest a positive correlation between service price and old-age dependency ratio in the long run.

The empirical part provides evidence for the positive correlation between the rise of the service price and old-age dependence ratio in the U.S. The regression results also show a positive correlation between service consumption and old households (age 65 and over) compared to the young. To bring these findings together, I then build up a small OLG model to suggest a potential mechanism behind these correlations. In this model, age affects service price through two channels: 1) on the demand side, labor markets with older populations tend to demand more services; 2) on the supply side, the education friction would cause worker mobility from the service sector to the manufacturing sector, which further increases the service price. The analytical results from the model confirm the results from price analysis, which means when the total population is stable, the more old generations tend to have a positive correlation between relative service price.

Chapter 2

Investigating Properties of Commodity Price Responses to Real and Nominal Shocks

2.1 Introduction

World commodity prices often exhibit highly persistent and volatile movements. As Deaton (1999) points out, correctly understanding the stochastic nature of commodity prices is crucial for enhancing the welfare of many developing countries that depend on the export of a few commodities. For example, if deviations of the commodity price from its equilibrium path are *short-lived*, the government may employ stabilization policies to mitigate the transitory impacts of the shock that caused the deviation. On the other hand, if the commodity price contains a unit root so that shocks result in a *permanent* change in the commodity price, policy-makers need to re-formulate their development strategies to incorporate such changes.

Early research in the commodity price literature has focused on the Prebisch-Singer hypothesis (PSH; Prebisch (1950), Singer (1950)). PSH implies a downward *deterministic* trend in the relative price of primary commodities to manufactured goods, continually deteriorating the terms of trade of those commodity-dependent countries. Sapsford (1985), Grilli and Yang (1988), and Helg (1991), among others, reported overall supportive evidence of PSH using commodity price *indices*, whereas Cuddington (1992), Bleaney and Greenaway (1993), and Newbold, Pfaffenzeller, and Rayner (2005) obtained very limited evidence using *disaggregated* commodity price data instead of using aggregate indices. More recently, Kellard and Wohar (2006), Harvey, Kellard, Madsen, andWohar (2010), and Ghoshray (2011)reported some nonlinear evidence in favor of PSH, allowing multiple structural breaks for a number of commodity prices.

A strand of researchers has estimated the persistence of commodity price shocks. For instance, Cashin, Liang, and McDermott (2000) claim that shocks to world commodity prices typically generate highly persistent effects. In a similar study, Cashin, McDermott, and Pattillo (2004) estimated bias-adjusted half-lives of the terms of trade shock for 42 sub-Saharan African countries. Although they reported *finite* half-life *point* estimates for majority (29 out of 42) countries, the point estimates were quite different across countries, ranging from 0.89-year to 34-year half life. Furthermore, most of their bias-corrected 90% confidence bands extended to positive infinity, meaning that statistical inferences on the length of the half-life are difficult due to high standard errors. Ghoshray (2013) also argued that the persistence of shocks varies widely across individual commodities and over time.

Researchers also have investigated the synchronization (comovement) of primary commodity prices. See, among others, Cashin, McDermott, and Scott (2002), Byrne, Fazio, and Fiess (2013), and West and Wong (2014). These comovement studies are closely related with an array of research works that analyze the source of underlying common driving forces/factors in the world commodity market. For example, Frankel (2008) highlighted the important role of the real interest rate in commodity price dynamics, while Chen, Rogoff, and Rossi (2010) point out that commodity currency exchange rates (e.g., Canadian dollar and Australian dollar) have substantial in-sample and out-of-sample predictive contents for commodity prices, but not in a reverse direction.

Another related research works estimate *latent common factors* applying the method of the principal component to a large panel of time series data. See, among others, Chen, Jackson, Kim, and Resiandini (2014), West and Wong (2014), Byrne, Fazio, and Fiess (2013). For instance, Chen, Jackson, Kim, and Resiandini (2014) demonstrated that the first common factor, estimated from a large panel of commodity price data, is closely related with the nominal exchange rate of the US dollar. Since these commodities are denominated in US dollars, their results confirm that the dollar exchange rate serves a common underlying driving force of world commodity prices.

In the present paper, we investigate statistical properties of price fluctuations in the world commodity market by estimating dynamic adjustment paths of the commodity price toward a new equilibrium path in response to unexpected changes in the nominal exchange rate and the world real GDP. We focus on these two primary factors to maintain a simple and homogeneous model structure for world commodity prices. Other potentially important factors such as storage costs, inventory levels, and short-term demand-supply conditions, see Williams and Wright (1991) and Deaton and Laroque (1992), are treated as idiosyncratic factors that are contained in the stationary error term. Instead of using commodity price indices or relative prices, we employ 49 dollar denominated world commodity prices, which are obtained from the IMF website, to make our empirical findings to be comparable with others in the current literature. For detailed explanations of these prices, see Table A1 in the Appendix.[1]

Using a vector autoregressive (VAR) model for the nominal exchange rate, the world real GDP, and the commodity price, we estimate the impulse-response function of 49 world

[1] Or check the IMF website at https://www.imf.org/external/np/res/commod/.

commodity prices in response to the exchange rate shock and the real GDP shock. We then define and estimate the dynamic elasticity of the commodity price with respect to these shocks. Instead of analyzing individual responses, we establish a number of stylized facts on commodity price dynamics utilizing kernel density estimates of the dynamic elasticity over time.

Our major findings are as follows. First, we observed the substantial degree of short-run *price stickiness* when the nominal exchange rate shock occurs. In the long-run, however, exchange rate changes are roughly absorbed by changes in the commodity price in dollars so that the commodity price stays constant in the rest of the world. That is, the *law of one price* (LOP) holds on average in the long-run, reflecting highly tradable nature of world commodities.

Second, the world real GDP shock (demand shock) tends to generate substantial price fluctuations on impact because adjustments of the supply can be quite limited in the short-run. The long-run elasticity with respect to the real GDP shock tends to be smaller than its short-run counterpart, because the supply can adjust fully to the shock and eventually counterbalances the increase in the demand in the long-run.

Third, we propose a measure of price stickiness. Kernel density estimates of this measure imply that the nominal exchange rate shock plays a more important role in explaining price dynamics in the long-run, whereas the real GDP shock contributes more to the short-run price dynamics. We also propose a measure of the contribution of the exchange rate shock relative to the real GDP shock, which confirms these findings. That is, nominal shocks in our empirical model have a more persistent long-lasting effect on commodity prices.

As Rogoff (1996) notes in his PPP puzzle, nominal shocks are considered to be short-lived, whereas real shocks yield slower adjustments toward the new equilibrium. On the contrary, Engel and Morley (2001) claimed that persistence of the real exchange rate is mainly driven by nominal shocks. Cheung, Lai, and Bergman (2004) also provided similar results. Our results are overall consistent with their findings.

The present paper also improves the work of Cashin, Liang, and McDermott (2000) and Cashin, McDermott, and Pattillo (2004) who used a univariate model that measures the persistence of the commodity price shock irrespective of the source of the shock. For example, we would expect a very different convergence path if unexpected changes in the commodity price was triggered by the exchange rate shock instead of the real GDP shock.

The rest of the paper is organized as follows. Section 2.2 presents our baseline VAR model framework. We also define the dynamic elasticity with respect to structural shocks. Section 2.3 provides a data description and reports our major empirical findings. Section 2.4 concludes.

2.2 The Econometric Model

We use a tri-variate vector autoregressive (VAR) model for the nominal exchange rate (e_t), the world real GDP (y_t), and the commodity price (p_t). All variables are log-transformed. p_t is ordered last in the VAR, meaning that other variables can influence it contemporaneously.[2]

Given p_t, unexpected increases in e_t (appreciations of the US dollar) result in higher commodity prices in the rest of the world ($e_t + p_t$). However, if p_t decreases sufficiently and offsets the increase in e_t, commodity prices in the rest of the world stay constant. When y_t rises unexpectedly, this serves as a positive demand shock in the commodity markets, resulting in an increase in p_t if the market supply fails to completely offset such an increase in the demand for commodities.

Since these variables are better approximated by an integrated process, that is, a nonstationary stochastic process, we employ VAR models after differencing the variables. Ab-

[2]Responses of p_t are robust to alternative ordering of e_t and y_t.

stracting from deterministic terms, we propose the following model.

$$x_t = \sum_{j=1}^{p} A_j x_{t-j} + C u_t, \tag{2.1}$$

where $x_t = [\Delta e_t, \Delta y_t, \Delta p_t]^{'}$, A_j denotes the j^{th} lag polynomial coefficient matrix, and C is the lower-triangular matrix that governs the contemporaneous relationship between the variables in x_t. $u_t = [u_t^e \ u_t^y \ u_t^p]'$ is a vector of mutually orthonormal structural shocks, that is, $Eu_t u_t^{'} = I$.

We obtain the orthogonalized impulse-response function (OIRF) for $\triangle e_t$ and $\triangle p_t$ to a one percent exchange rate shock u_t^e as follows. $\rho_e^p(j) = E\left(\Delta p_{t+j} | u_{e,t} = 1, I_{t-1}\right) - E\left(\Delta p_{t+j} | I_{t-1}\right)$ $\rho_e^e(j) = E\left(\Delta e_{t+j} | u_{e,t} = 1, I_{t-1}\right) - E\left(\Delta e_{t+j} | I_{t-1}\right)$, where I_{t-1} is the adaptive information set at time $t-1$. Response functions of the *level* variables are obtained by cumulatively summing these response functions.

$$\psi_e^p(j) = \sum_{s=0}^{j} \rho_e^p(j), \ \psi_e^e(j) = \sum_{s=0}^{j} \rho_e^e(j), \tag{2.2}$$

that is, $\psi_e^p(j) = E(p_{t+j} | u_{e,t} = 1, I_{t-1})$ and $\psi_e^e(j) = E(e_{t+j} | u_{e,t} = 1, I_{t-1})$, because $p_{t-1} = e_{t-1} = 0$.[3]

p_t and e_t are log-transformed series, therefore $\psi_e^p(j)$ and $\psi_e^e(j)$ are expected growth rates of the commodity price and the exchange rate over j period when the exchange rate shock occurs at time t.[4] We define the following dynamic elasticity of the commodity price at time $t+j$ with respect to the exchange rate as follows.

$$\eta_e^p(j) = \frac{\psi_e^p(j)}{\psi_e^e(j)} \tag{2.3}$$

Similarly, we define the dynamic elasticity of a commodity price with respect to the real GDP.

$$\eta_y^p(j) = \frac{\psi_y^p(j)}{\psi_y^y(j)}, \tag{2.4}$$

[3]Recall that x_t is demeaned prior to estimations.
[4]That is, $\ln Z_{t+j} - \ln Z_t \approx (Z_{t+j} - Z_t)/Z_t$.

where $\psi_y^p(j)$ and $\psi_y^y(j)$ are the response function of the level variables p_t and y_t at time $t+j$, respectively, when there is a shock to the real GDP. $\eta_e^p(0)$ and $\eta_y^p(0)$ denote the contemporaneous dynamic elasticity, while $\eta_e^p(\infty)$ and $\eta_y^p(\infty)$ are the long-run dynamic elasticity of the commodity price.

We also propose a measure of stickiness of the commodity price as follows.

$$\kappa_e = \frac{\psi_e^p(0)}{\psi_e^p(\infty)},\ \kappa_y = \frac{\psi_y^p(0)}{\psi_y^p(\infty)} \tag{2.5}$$

For example, κ_e is the share of the initial response of the commodity price to the exchange rate shock relative to its long-run response, whereas κ_y is a similarly defined measure when there is a real GDP shock. Note that these measures provide information on price rigidity when each of these shocks occur. A small positive κ_e or κ_y implies a higher degree of price rigidity, whereas high positive values mean that price adjustments mostly take place on the impact of the shock. A negative number implies that the sign of the response changes over time, which often comes with a wide confidence band that implies an insignificant response.

Lastly, we define the following index to measure the contribution of the exchange rate shock for the j-period ahead forecast variations in the commodity price relative to the world real GDP (demand) shock.

$$\phi(j) = \frac{|\eta_e^p(j)|}{|\eta_e^p(j)| + |\eta_y^p(j)|} \tag{2.6}$$

Naturally, the relative contribution of the world demand shock is defined by $1-\phi(j)$.

In what follows, we employ the following nonparametric kernel density estimator for the density function of the random variable $x = \eta_e^p(j),\ \eta_y^p(j),\ \kappa_e,\ \kappa_y,\ \phi(j)$.

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{X_i - x}{h}\right), \tag{2.7}$$

where n is the number of commodity prices, h is the bandwidth parameter, X_i is the observation of the random variable x, and $k(\cdot)$ denotes a kernel function.[5] We choose the optimal h by

[5] We employ the Epanechnikov kernel and Gaussian kernel, which yield similar results.

conventional Silverman's rule of thumb.

2.3 Data Descriptions and Empirical Findings

2.3.1 Data Descriptions

We obtained 49 primary world commodity prices (p_t) from the International Monetary Fund (IMF) website. The data set includes 23 food prices (7 cereals, 5 vegetable oils, 4 meats, 3 seafoods, and 4 other foods), 4 beverage prices, 9 agricultural raw material prices, 8 metal prices, and 5 fuel prices. For details, see Table A1 in the appendix. All commodity prices are denominated in the US dollar. We transformed original monthly frequency commodity prices to quarterly frequency series by taking the end of period value, because the world real GDP growth rate (Δy_t; 00199BPXZF), obtained from the International Financial Statistics (IFS) CD-ROM, is available in quarterly frequency. Observations span from 1980:I to 2013:IV.

The nominal exchange rate (e_t) is the trade-weighted average US dollar index for major currencies (TWEXMMTH) that include the Euro area, Canada, Japan, United Kingdom, Switzerland, Australia, and Sweden. We obtained the monthly frequency data from the Federal Reserve Economic Data (FRED) for the same sample period, then transformed it to quarterly data.

2.3.2 Empirical Findings

We first estimate the tri-variate VAR model in (2.1) for each commodity price (p_t), then obtain the orthogonalized cumulative impulse-response function estimates as defined in (2.2) and (2.2).

It should be noted that responses of the nominal exchange rate (e_t) and the real GDP (y_t) to their *own* shock are quantitatively very similar no matter what p_t is used in (2.1).

On average, e_t increases by 1.08% in the long-run in response to $u_t^e = 1\%$. The standard deviation of the responses was 0.02%, which implies a very tight distribution of the estimate across commodities. The average response of y_t in the long-run was 4.60% when there is a one percent shock to u_t^y. The distribution is again very tight with 0.18% standard deviation. That is, we obtained robust estimates for $\psi_e^e(\infty)$ and $\psi_y^y(\infty)$. Initial responses, $\psi_e^e(0)$ and $\psi_y^y(0)$, were also quantitatively very similar across commodities.[6] On the other hand, responses of commodity prices (p_t) to u_t^e and u_t^y, that is, $\psi_e^p(\cdot)$ and $\psi_y^p(\cdot)$ exhibit a high degree of heterogeneity across commodities, which will be discussed in what follows.

Figure 2.1 reports some example impulse-response function estimates for corn (PMAIZMT) and Brent oil (POILBRE) prices along with their associated 95% confidence intervals that are obtained from 500 nonparametric bootstrap simulations. Corn price decreases by 0.76% on impact when one standard deviation exchange rate shock ($u_t^e = 3.63\%$) occurs, whereas Brent oil price decreases by 4.24% when the same shock occurs. In terms of the dynamic elasticity, these responses correspond to -0.21 and -1.17 for corn and Brent oil prices, respectively. That is, corn price exhibits a contemporaneously inelastic response, which implies a substantial degree of short-run price stickiness. On the other hand, Brent oil price slightly over-corrects (more than one-for-one adjustment) the exchange rate shock in the short-run. The long-run elasticity estimates are -1.23 for corn price and -0.99 for Brent oil price, respectively.[7,8] That is, corn price over-corrects the exchange rate shock, while Brent oil price just-corrects it in the long-run.

In response to a one standard deviation real GDP shock ($u_t^y = 1.58\%$), corn and Brent oil prices increase by 0.44% and 2.41% on impact, while they rise around by 7.04% and 5.34% in the long-run, respectively. The corresponding dynamic elasticity estimates for corn price

[6] All results are available upon request.

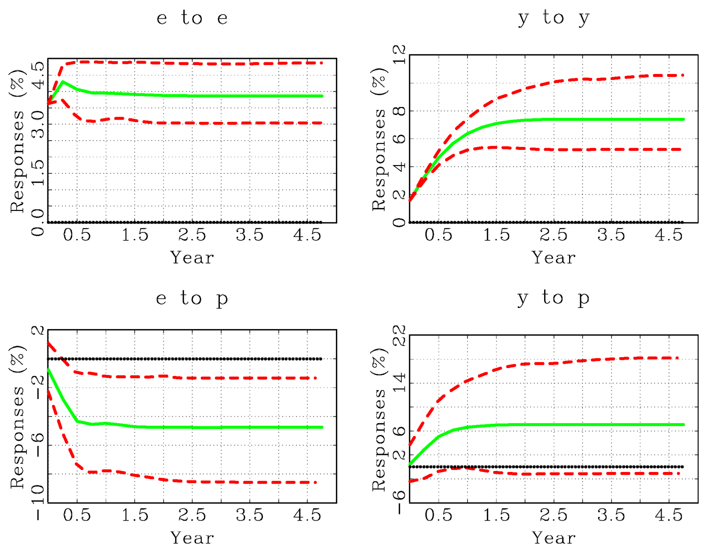
[7] We estimate the long-run elasticity by taking the elasticity estimate for $j = 40$ (10 years) which is long enough for the responses to converge.

[8] See Tables 2 and 3 for dynamic elasticity estimates for all 49 commodity prices with respect to the exchange rate shock.

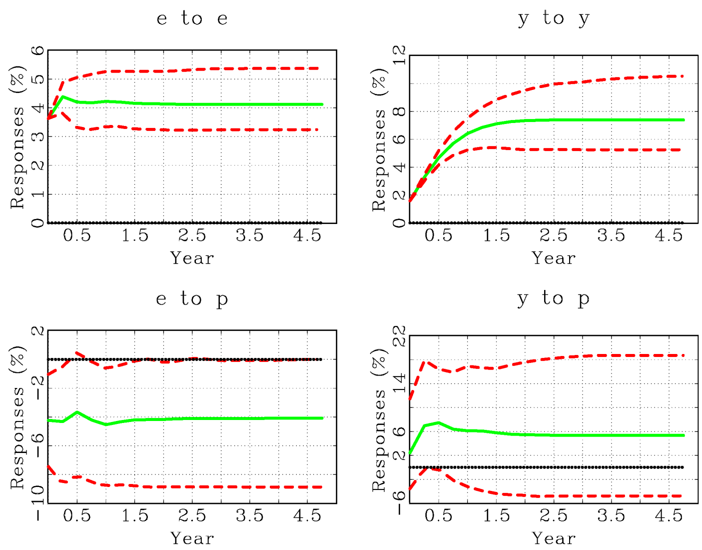
Figure 2.1: Examples of the Impulse-Response Function Estimates to One Standard Deviation Shocks

$$\psi_e^e(j) = \sum_{s=0}^{j} \rho_e^e(s),\ \psi_e^p(j) = \sum_{s=0}^{j} \rho_e^p(s),\ \psi_y^y(j) = \sum_{s=0}^{j} \rho_y^y(s),\ \psi_y^p(j) = \sum_{s=0}^{j} \rho_y^p(s)$$

(a) Corn



(b) Brent



Note: The magnitude of the shock is one standard deviation of each variable, 3.634% and 1.581% for the exchange rate return and the world real GDP growth rate, respectively. Point estimates (solid lines) are reported with 95% confidence bands (dashed lines) that are obtained by 500 nonparametric bootstrap simulations.

are 0.28 and 0.95 in the short-run and in the long-run, respectively. On the other hand, the dynamic elasticity of Brent oil price are 1.52 in the short-run and 0.72 in the long-run.[9] Note that Brent oil price over-reacts to the real GDP shock in the short-run, but its long-run response is somewhat muted.

In what follows, we establish a number of stylized facts on world commodity price responses to the nominal exchange rate and the real GDP shocks based on empirical distributions of the dynamic elasticity estimates.

Figure 2.2 reports kernel density estimates of the 49 dynamic elasticity point estimates with respect to the exchange rate, $\eta_e^p(j) = \frac{\psi_e^p(j)}{\psi_e^e(j)}$. We consider the contemporaneous elasticity ($j = 0$) as well as the long-run elasticity ($j = \infty$).[10] We also report the point estimate for each of 49 commodity prices as well as its percentiles, $p_{0.05}$, $p_{0.50}$, and $p_{0.95}$ (p_x is the x percentile), that are obtained from 500 nonparametric bootstrap simulations in Tables A2 and A3 in the appendix. Note that $p_{0.05}$ and $p_{0.95}$ constitute the 90% nonparametric confidence band for each commodity price.

The median (mean) value of the contemporaneous elasticity, $\eta_e^p(0)$, was -0.66 (-0.59), while those of the long-run elasticity, $\eta_e^p(\infty)$, was -0.94 (-0.94). It should be noted that $\eta_e^p(j) = -1$ implies that changes in the exchange rate (e_t) are completely absorbed by changes in the commodity price (p_t). That is, the commodity price stays constant in terms of the rest of the world price ($p_t^* = e_t + p_t$), which is consistent with the law of one price (LOP) proposition. Naturally, we choose $\eta_e^p(j) = -1$ as a benchmark for a just-correction case. Given that, the median (or mean) of $\eta_e^p(0)$ implies a sluggish price adjustment in the short-run, whereas the median (or mean) of $\eta_e^p(\infty)$ is roughly consistent with LOP in the long-run.

To statistically evaluate the possibility of price-stickiness, we implemented a two-sided

[9] See Tables 4 and 5 for dynamic elasticity estimates for all 49 commodity prices with respect to the real GDP shock.

[10] Again, we estimate the long-run elasticity by taking the elasticity estimate for $j = 40$, which is long enough for the responses to converge.

t-test with the null hypothesis of zero degree of price-stickiness, $H_0 : \eta_e^p(j) = -1$. The test rejects the null hypothesis for the contemporaneous ($j = 0$) elasticity at the 1% significance level ($t = 5.84$), while it fails to reject the null for the long-run ($j = \infty$) elasticity at any conventional significance level ($t = 0.86$). That is, we obtained very strong evidence of short-run price rigidity. In the long run, the elasticity estimates are centered around the benchmark value (-1), which means that commodity prices, on average, counter-balances the effect of the exchange rate shock in the long-run.

These findings overall imply that LOP holds on average in the world commodity market, even though there exists a non-negligible degree of heterogeneity across individual commodities. The kernel density estimates are fairly wide both in the short-run and in the long-run. As we can see in Tables A1 and A2 in the appendix, cereal prices (#1 to #7) tend to exhibit a muted contemporaneous response to the exchange rate shock, while oil prices (#46 to #49) instantly adjust to the benchmark value when the shock occurs, which implies a greater degree of efficiency in crude oil markets.

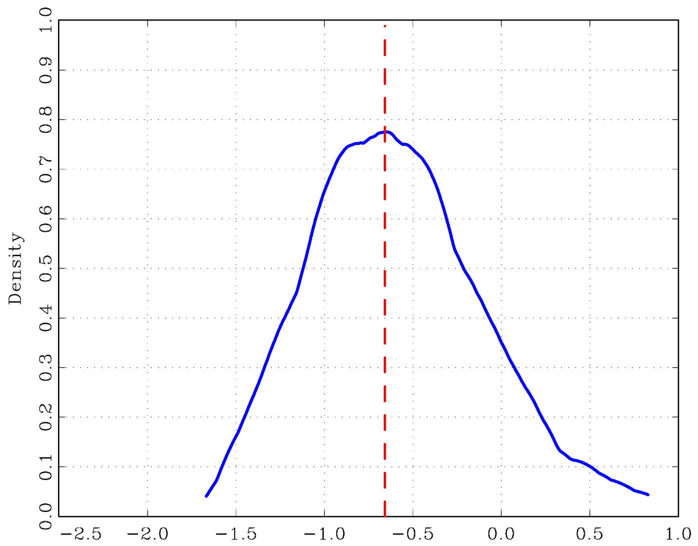
In Figure 2.3, we report kernel density estimates of the 49 dynamic elasticity point estimates with respect to the real GDP, $\eta_y^p(j) = \frac{\psi_y^p(j)}{\psi_y^y(j)}$. The median (mean) is 0.89 (0.85) and 0.25 (0.23) for $\eta_y^p(0)$ and $\eta_y^p(\infty)$, respectively. Complete results for individual prices are reported in Tables A4 and A5 in the appendix. We select $\eta_y^p(j) = 0$ as a benchmark elasticity, which may happen when the real GDP (demand) shock is completely absorbed by corresponding changes in the supply of the commodity.

As we can see in Figure 2.3, dynamic elasticity tends to be greater in the short-run than in the long-run. This means that commodity markets tend to rely on price adjustment in the short-run when there's a positive real GDP shock (demand shock), because short-run quantity adjustments in the supply tend to be limited. On the other hand, positive demand shocks seem to greatly promote the supply of commodities in the long-run, which then curb further rises in

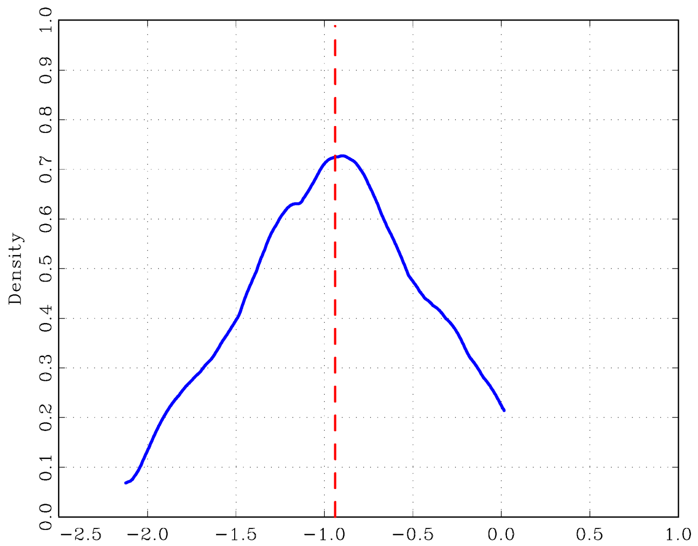
Figure 2.2: Kernel Density Estimations of the Dynamic Elasticity: Exchange Rate Shock

$$\eta_e^p(j) = \frac{\psi_e^p(j)}{\psi_e^e(j)},\ \psi_e^p(j) = \sum_{s=0}^{j} \rho_e^p(s),\ \psi_e^e(j) = \sum_{s=0}^{j} \rho_e^e(s)$$

(a) Contemporaneous Elasticity $\eta_e^p(0)$



(b) Long-Run Elasticity $\eta_e^p(\infty)$



Note: We use the Epanechnikov kernel to estimate the kernel density functions. The vertical dashed line is the median value of the point estimate, -0.658 and -0.939 for $\eta_e^p(0)$ and $\eta_e^p(\infty)$, respectively. The t-statistic for the null hypothesis of zero price-stickiness (-1) was 5.841 and 0.855 for $\eta_e^p(0)$ and $\eta_e^p(\infty)$, respectively. That is, the test strongly supports the short-run price rigidity, whereas the null is accepted for the long-run elasticity.

the commodity price. Consequently, the long-run dynamic elasticity tends to be smaller than the short-run elasticity when there's a real GDP shock.

Recall that the exactly opposite was true when exchange rate shocks occur. That is, these findings provide empirical evidence that nominal shocks can have more *pronounced* long-lasting effects on the commodity price than real shocks, which is consistent with the findings of [32] and [21].

We also note that the standard deviation of the long-run elasticity (0.64) is much smaller than that of the short-run elasticity (1.64), which implies a greater degree of heterogeneity of the short-run responses than the long-run responses to the real GDP shock. One interesting finding (See Tables A1 and A4 in the appendix) is that such a short-run heterogeneity is mainly driven by short-run elasticities of food and beverage (necessities) prices (#1 to #27). The average short-run elasticity of these commodities is 0.38, which is far smaller than that of agricultural raw materials, metals, and fuel prices (#28 to #49), 1.42. On the other hand, both groups exhibit about the same average long-run elasticity around 0.2.

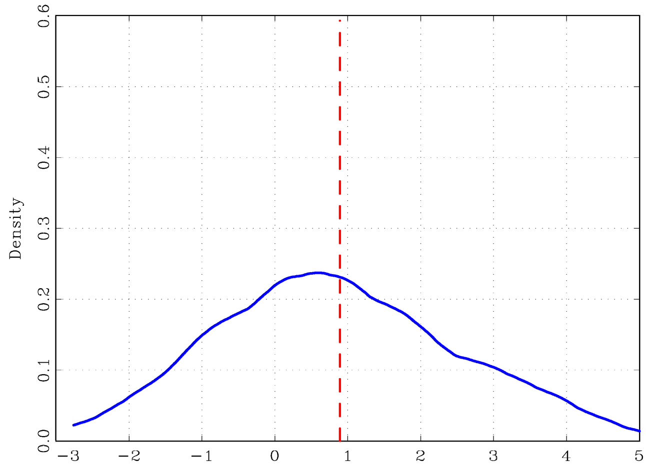
Again, we implement a two-sided t-test with the null hypothesis, $H_0 : \eta_y^p(j) = 0$. The t-statistic was 3.65 and 2.49 in the short-run and in the long-run, respectively. Even though the test rejects the null hypothesis for both cases, the t-statistic is greater (smaller p-value) for the short-run elasticity, meaning that the test provides a stronger evidence against the null hypothesis in the short-run.

Figure 2.4 presents kernel density estimates of the price rigidity measure in (2.5), $\kappa_e = \psi_e^p(0)/\psi_e^p(\infty)$ and $\kappa_y = \psi_y^p(0)/\psi_y^p(\infty)$. We first note that most κ_e estimates are positive (43 out of 49) and are distributed around its median value 0.64. Ruling out obvious outliers, the estimated distribution is quite compact and supports a partial adjustment ($\kappa_e < 1$) in the short run. Put it differently, we report substantial degree of sluggish adjustments of commodity prices when there is an exchange rate shock.

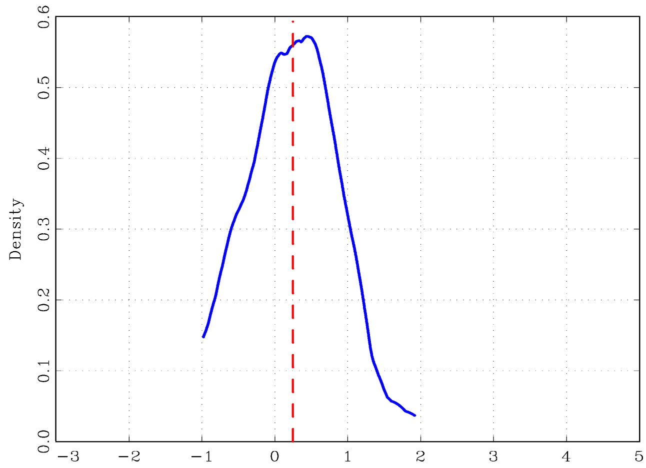
Figure 2.3: Kernel Density Estimations of the Dynamic Elasticity: Real GDP Shock

$$\eta_y^p(j) = \frac{\psi_y^p(j)}{\psi_y^y(j)},\ \psi_y^p(j) = \sum_{s=0}^{j} \rho_y^p(s),\ \psi_y^y(j) = \sum_{s=0}^{j} \rho_y^y(s)$$

(a) Contemporaneous Elasticity $\eta_y^p(0)$



(b) Long-Run Elasticity $\eta_y^p(\infty)$



Note: We use the Epanechnikov kernel to estimate the kernel density functions. The vertical dashed line is the median value of the point estimate, 0.855 amd 0.229 for $\eta_y^p(0)$ and $\eta_y^p(\infty)$, respectively The standard deviation was 1.637 and 0.640 for $\eta_y^p(0)$ and $\eta_y^p(\infty)$, respectively, indicating more homogeneous responses across commodities in the long-run. The t-statistic for the null hypothesis of no effect (0) was 3.648 and 2.485 for $\eta_y^p(0)$ and $\eta_y^p(\infty)$, respectively. That is, the test implies a stronger effect of the demand shock in the short-run than in the long-run, even though the test rejects the null in both cases.

On the other hand, κ_y estimates are widely distributed around its median 0.30 with a large standard deviation (1.42). Large κ_y estimates in *absolute value* imply that prices fluctuate substantially in the short-run when real GDP shocks occur, whereas the impacts of the real GDP shock becomes muted in the long-run possibly due to sufficiently large adjustments of the supply of commodities that counterbalance the increase in the demand. Note that results in Figure 2.4 are overall consistent with our interpretations on results in Figures 2.2 and 2.3.

In a nutshell, these density estimates imply that the nominal exchange rate shock plays a more important role in explaining commodity price dynamics in the long-run relative to the real GDP shock, which contributes more to short-run dynamics of commodity prices.

We further investigate these properties in depth by estimating the kernel density of the relative contribution of the exchange rate using the index in (2.6), $\phi(j) = |\eta_e^p(j)| \,/\, \left[|\eta_e^p(j)| + \left|\eta_y^p(j)\right|\right]$. See Figure 5. The median (mean) $\phi(j)$ estimate is 0.35 (0.38) contemporaneously ($j = 0$), while the median (mean) increases to 0.62 (0.62) in the long run.[11] That is, these estimates imply that the exchange rate shock contributes more to long-run price dynamics, whereas the real GDP (demand) shock influences the commodity price more dominantly in the short-run.

These findings are again consistent with our previous empirical results. Nominal exchange rate shocks have limited effects on commodity prices in the short-run exhibiting price rigidity, whereas commodity prices fluctuate greatly on impact when real GDP shocks occur, because adjustments in the supply of commodities can be sluggish in the short-run.

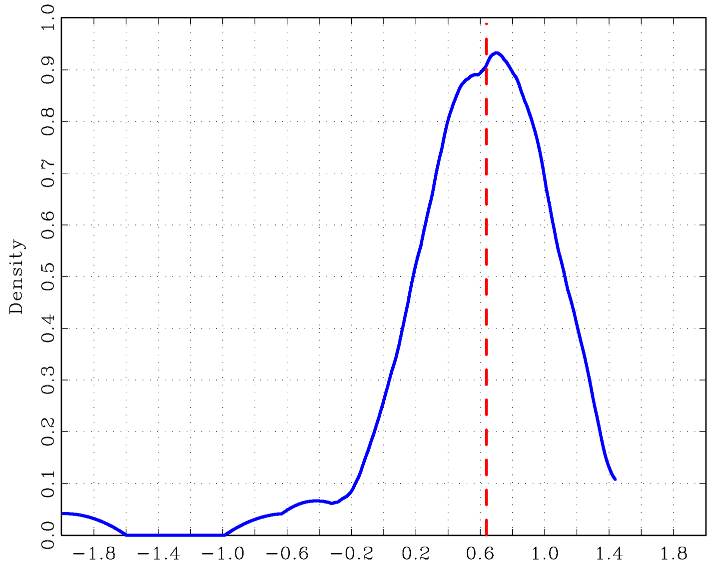
In the long run, on the other hand, LOP forces world commodity prices to respond more substantially to changes in the exchange rate via commodity arbitrages. On the other hand, the effect of the real GDP shock becomes weak as adjustments in the supply of commodities curb the influence of increases in the world real GDP.

[11] We report full reports in Tables A6 and A7 in the appendix.

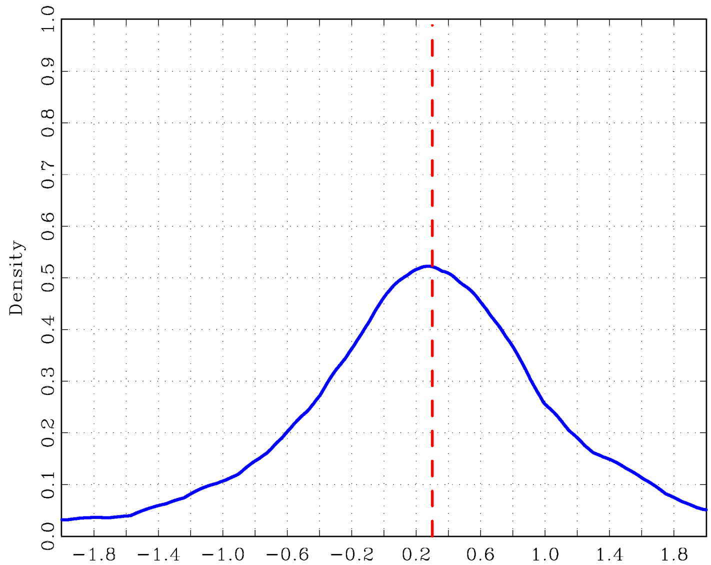
Figure 2.4: Kernel Density Estimations of the Price Rigidity Measure

$$\kappa_e = \frac{\psi_e^p(0)}{\psi_e^p(\infty)},\ \kappa_y = \frac{\psi_y^p(0)}{\psi_y^p(\infty)},\ \psi_e^p(j) = \sum_{s=0}^{j} \rho_e^p(s),\ \psi_y^p(j) = \sum_{s=0}^{j} \rho_y^p(s)$$

(a) Exchange Rate Shock κ_e



(b) Real GDP Shock κ_y

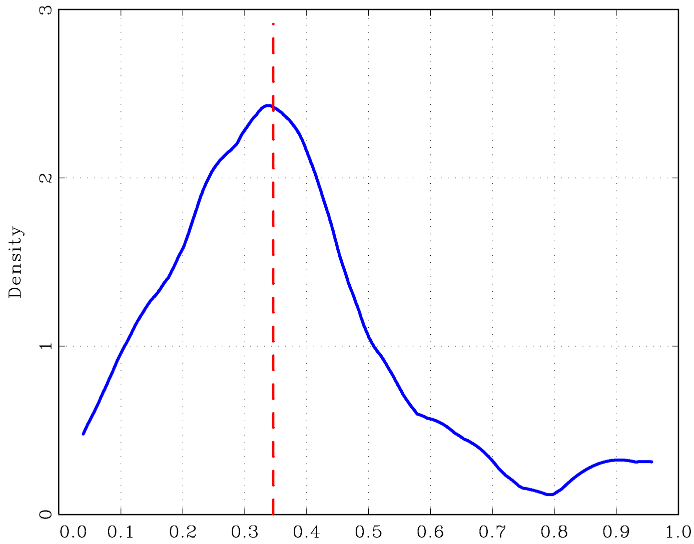


Note: We use the Epanechnikov kernel to estimate distributions. The vertical dashed line is the median value of the point estimate, 0.638 and 0.299 for κ_e and κ_y. The standard deviation was 0.548 and 1.423 for κ_e and κ_y. We obtained strictly positive κ_e estimates for 43 out of 49 commodities, whereas κ_y estimateis were positive only for 32 commodities.

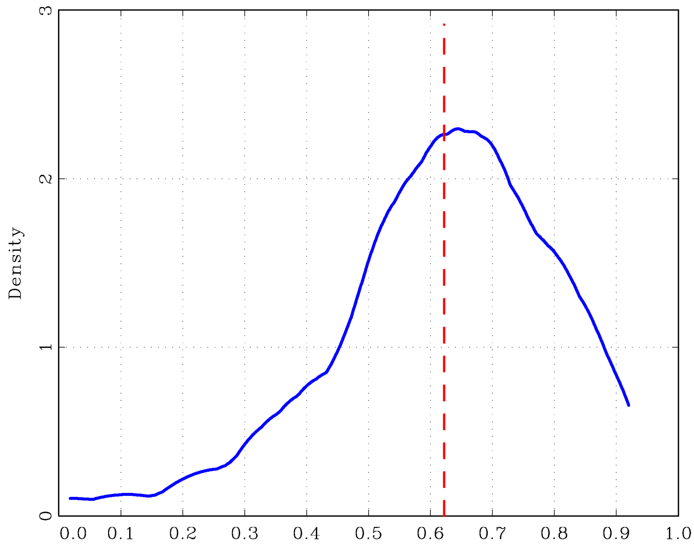
Figure 2.5: Kernel Density Estimations of the Relative Dynamic Elasticity

$$\phi(j) = \frac{|\eta_e^p(j)|}{|\eta_e^p(j)| + \left|\eta_y^p(j)\right|}$$

(a) Contemporaneous Relative Elasticity $\phi(0)$



(b) Long-Run Relative Elasticity $\phi(\infty)$



Note: We use the Epanechnikov kernel to estimate distributions. The vertical dashed line is the median value of the point estimate, 0.346 and 0.622 for $\phi(\infty)$and $\phi(\infty)$, respectively. That is, the exchange rate shock plays a more important role relative to the real GDP shock in the long-run, while the opposite is true in the short-run.

Lastly, we repeat kernel density function estimations using real commodity prices as a robustness check analysis. For this, we deflated all commodity prices using the US consumer price index (CPI) because all commodities are denominated in the US dollar. We obtained quantitatively very similar results, which is not surprising because dynamics of nominal commodity prices are similar to real prices because the CPI exhibits much less variations compared with individual commodity prices. All results are reported in Figure 2.6.

2.4 Concluding Remarks

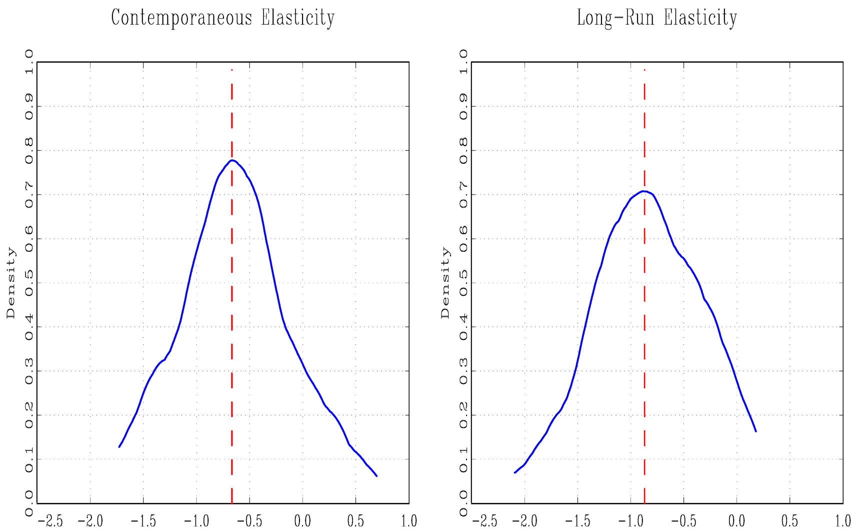
This paper estimates and compares dynamic responses of 49 world commodity prices to unexpected changes in the nominal exchange rate and the world real GDP growth rate. Instead of looking at individual responses, we utilize kernel density function analysis to establish a number of stylized facts on commodity price adjustments toward a new equilibrium after these shocks occur. Our major findings are as follows.

First, we report strong evidence of short-run price rigidity in the world commodity market when nominal exchange rate shocks occur. However, changes in the exchange rate, on average, are absorbed by corresponding changes in commodity prices in the long-run so that the commodity price stays constant in the rest of the world. That is, the law of one price holds in the long-run.

Second, the world real GDP shock has a substantial positive effect on the commodity price in the short-run. On average, the commodity price increases by over 0.8% in the short-run when there's a 1% shock. However, we obtained a fairly flat kernel density function that implies a high degree of heterogeneity across international commodity markets. On the other hand, the real GDP shock has a very weak impact on commodity prices in the long-run, as the supply of commodities eventually counterbalances the changes in the demand triggered by the real GDP

Figure 2.6: Kernel Density Estimation Results with Real Commodity Prices

(a) Exchange Rate Shock



(b) Real GDP Shock

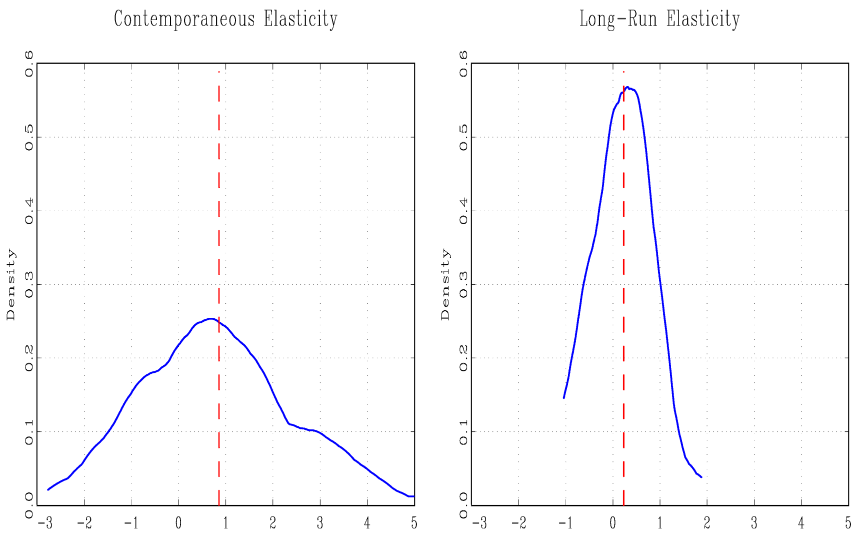
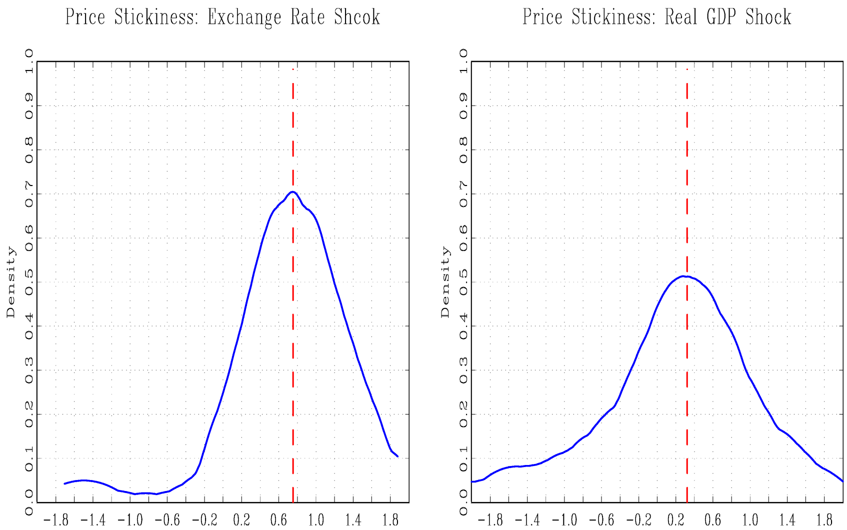
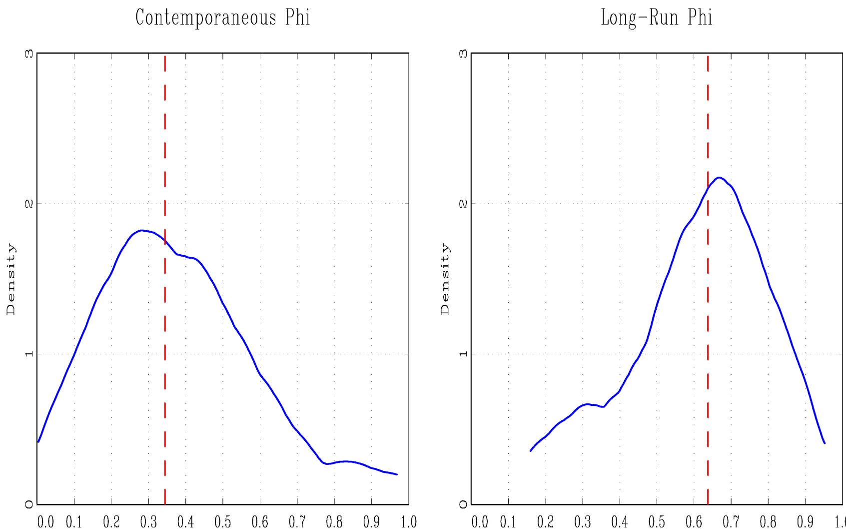


Figure 2.7: Continued

(c) Price Stickiness Measure



(d) Relative Dynamic Elasticity



Note: The point estimate distribution is solid line. We use the Epanechnikov kernel to estimate distributions. The vertical dashed line is the median value of the point estimate.

shocks in the long-run.

Third, we propose a measure of price rigidity, which is a share of the short-run response of the commodity price relative to its long-run response. Our kernel density analysis implies a high degree of price stickiness when the exchange rate shock occurs. In response to the real GDP shock, we find much weaker and heterogeneous evidence of price rigidity across commodities.

Lastly, we define and estimate the contribution index of the nominal exchange rate shock relative to the real GDP shock to fluctuations in commodity prices. Our results imply that the nominal exchange rate plays relatively more important role in explaining commodity price dynamics in the long-run, whereas the real GDP shock contributes more to short-run price fluctuations.

Chapter 3

News Sentiment and Topic Analysis on Crude Oil Future Prices

3.1 Introduction

Crude oil is a commodity, and as such, it tends to have large fluctuations at price than more stable investments such as stocks and bonds. Crude oil prices are influenced by a variety of factors. As Baumeister and Kilian (2014) conclude that the explanatory power of these factors vary over time and that different factors are important at different time horizons. Though the factors to explain the change of oil prices are inconclusive, Brandt and Gao (2019) mention that news information can provide a way to quantify macroeconomic and other events that could affect crude oil prices. In this paper, we analyze the contents of news articles to study how information about crude oil related news affects crude oil futures price.

News information could affect crude oil prices in different ways. Shiller (2015) argues that the news media plays an important role in setting the stage for market moves and provoking them. On one hand, news could convey information on the current significant variables that affect the price, reflecting the current market confidence. On the other hand, news could serve as an update of the changes in these significant variables, reflecting the expectations about future oil supply and demand conditions, which could affect the current crude oil prices.

In this paper, we consider broad web news from various sources. We analyze these web crude oil related news to uncover how this information affects the crude oil futures price using the intra-day high-frequency data. To do so, we use both supervised and unsupervised machine learning algorithms to study the impact of news sentiment and news topics on crude oil futures price increase. The important assumption in our analysis is that the crude oil futures market is nearly perfectly efficient, which means that the price can adjust quickly to any new information released to the public.

A key contribution of our paper is to demonstrate finer and more objective classifications of news effects on crude oil price change. First, we use unsupervised machine learning algorithms to group crude oil related news articles into different topics without providing prior knowledge on how each topic links to a particular set of words. Second, we conduct both news sentiment analysis and news topics analysis of each news article using intra-day high- frequency data, these results providing us a new index data indicating crude oil price increase or decrease for future studies based on textual analysis.

In this paper, first, we conduct news sentiment analysis with logistic regression to see how each word in an article can affect the increase or decrease of the crude oil price from September 2019 to December 2019. We find that among all the news words, 152 words have coefficients smaller than -0.5, which are collected as the most negative words. Also, 159 words have coefficient over 0.5, we defined these words as the most positive words. Based on the news sentiment analysis results, we also construct a novel index indicating the sentiment score of a news article given the most positive (negative) words. Second, we categorize news articles discussing crude oil over the entire year of 2019 into topics using unsupervised machine learning algorithms K-means. The K-means algorithm generates 4 clusters of news topics. We rename and interpret each of them to be “World Crude Oil” topic, “WTI Crude Oil” topic, “Financial Analysis” topic, and “Editorial Opinion” topic. Each of the news articles would be assigned to be one of the topics. Finally, we estimate how the news topic and sentiment score would affect the crude oil future price using logistic regression in 5 minutes window. The results suggest that on average, ”World Crude Oil” news has the highest correlation with a crude oil price increase. Moreover, the more positive news is under the topic ”WTI Crude Oil,” the higher probability that WTI crude oil futures price will increase within 5 minutes.

The remainder of this paper is organized as follows: Section 3.2 presents the related literature. Section 3.3 provides detailed data information. Section 3.4 shows the news sentiment analysis methodology and how the positive score is calculated for every news article. Section 3.5 details the construction of topic analysis. Section 3.6 presents our main empirical analysis. Section 3.7 concludes.

3.2 Literature Review

This paper mainly relates to three streams of research: (1) studies on general crude oil price; (2) studies on news effects of crude oil price; (3) studies on news textual analysis on general equity market. Broad literature has studied the explaining factor of crude oil price. Hamilton (2009) suggests that the real price of oil follows a random walk without drift. Rapaport (2013) distinguish between demand and supply driven component of crude oil returns by examining its correlation with the equity market. Baumeister and Kilian (2014) discuss an exhaustive set of oil pricing factors from the literature, conclude that the explanatory power of these factors varies over time and that different factors are important at different time horizons. These findings crucially depend on the underlying model structure and assumptions. This paper corroborates these effects from the assumption that the crude oil futures market is nearly perfectly efficient, updated news information is a direct way to reflect the market sentiment.

Our paper is part of a growing body of research using textual analysis to examine how news affects economic and financial variables. Most work in this literature deals with the general equity market and aggregate news about equities and the economy. For example, Garcia (2013) studies the effect of sentiment on asset prices using the New York Times between 1905 and 2005. They find that after controlling for other well-known time-series patterns, the predictability of stock returns using news' content is concentrated in recessions. Soo (2015) develops a measure of sentiment across local housing markets by quantifying the positive and negative tone of housing news in local newspaper articles. Bi and Traum (2020) examines how newspaper reporting affects government bond prices during the U.S. state default of the 1840s. Our paper is different from the aforementioned studies in that we consider a much broader set of news resources and news categories as inputs, we can capture both micro-level supply and demand factors and macroeconomic related factors, even geopolitical developments.

This paper also contributes to the literature about news effects on the crude oil prices. Most of the literature work with the effects of regularly scheduled macroeconomic releases on crude oil price. For example, Kilian and Vega (2011) propose a formal test of the identifying assumption that energy prices are predetermined for U.S. macroeconomic aggregates using daily energy prices on daily news from the U.S. macroeconomic data releases. In this aspect, this paper is different from the existing literature focusing on responses of crude oil prices to scheduled macro news announcement from U.S. Another difference from papers on the schedule announcement is that our news indices are at high frequency than the scheduled releases with a fixed frequency. A recent paper by Brandt and Gao (2019) uses sentiment scores for global news from RavenPack global macro package to see how news about macroeconomic fundamentals and geopolitical events affect crude oil markets. Our paper is different from this paper since we use machine learning algorithms to group news into different topics, which avoid prior knowledge and generate more objective classifications of news effects on crude oil price change.

3.3 Data

3.3.1 Data Source and Description

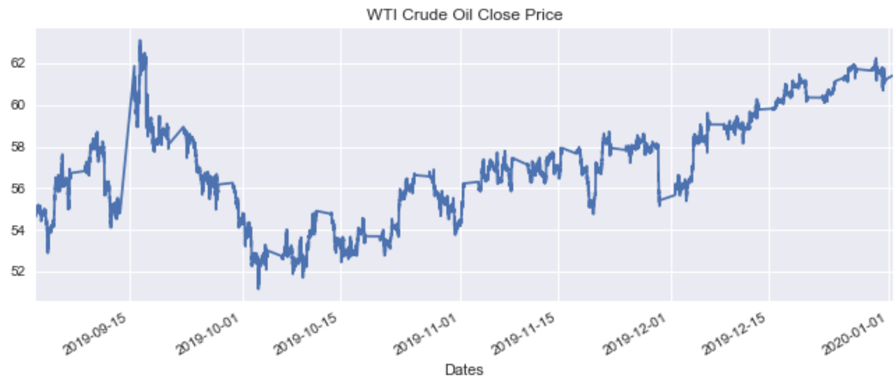
The data used in this paper are obtained through Bloomberg terminal, including the west texas intermediate (WTI) crude oil futures prices, and crude oil-related news articles. This paper focuses on WTI crude oil. WTI refers to oil extracted from wells in the US and sent via pipeline to Cushing, Oklahoma. This paper chooses WTI crude oil over Brent crude oil, which counts two-thirds of all crude contracts around the world, and Dubai crude oil, which is the major supply for the Asian market. The reason is that WTI crude oil has been the main benchmark for crude oil consumed in the United States. Thus its price is more related to the supply and demand conditions in the US market than in other markets worldwide.

The intra-day trading data is provided by Bloomberg, with price updating at high frequencies. However, the trading will be suspended every day at 14:00 to 15:00, as well as every Saturday. In order to match the frequency of news, this paper uses high-frequency intra-day trading futures prices for WTI crude oil. The price is updating every five minutes, which matches the first quantile of news frequency distributions[1]. This paper takes the sample period to be the last quarter of 2019. Figure 3.1 shows the WTI crude oil close price throughout the period. From Figure 3.1, it is clear that the close price has high volatility but no obvious trend during the last quarter of 2019, which makes it a perfect sample for this paper's analysis.

News articles are the major sources for text analysis. This paper analyzes contents

[1]Shown in Figure 3.4

Figure 3.1: Close Price Fluctuation in Sample Period



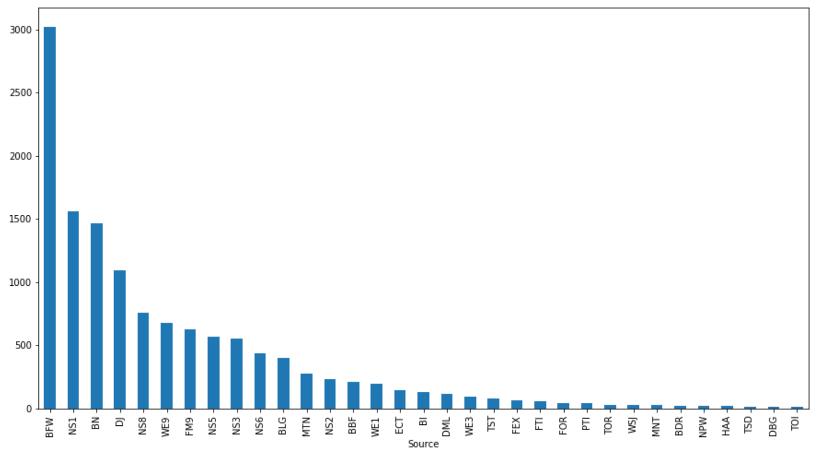
of each news article to find the news effect on WTI crude oil futures price. The news articles are collected through the Bloomberg terminal's news section. By specifying the news topics to be 'Crude Oil', including the keyword, 'WTI', Bloomberg terminal gives editorial recommended Bloomberg and web news from various sources. Figure 3.2 is a word cloud showing the most frequent words in the news articles collected by this paper. In 2019, news articles under the topic of WTI crude oil discussed the 'oil price', 'trade war', 'saudi arabia', 'texax intermediate', etc. More detail analysis in news contents will be presented in Section 3.4.

Figure 3.3 shows the major news sources for this paper's dataset. In total, there are

Figure 3.2: 2019 WTI Crude Oil Related News Frequent Words

75 news sources. Figure 3.3 plots the news sources with total news article released counts over ten. All news are global news written in English, reporting crude oil-related topics worldwide. Around 6000 news articles are web news, Bloomberg provides the news subject and the link of the websites. Web news content is collected through web scraping. Two-thirds of the web news is successfully obtained through web-scraping, with the rest replaced by its subject. Throughout 2019, there are 13183 news articles, and 4615 of them are released after September. Since this paper only has price data from September 2019 to December 2019, only news articles released after September will be included in news sentiment analysis and the final regressions, but all news articles in 2019 will be included in news topic analysis.

Figure 3.3: Major News Source

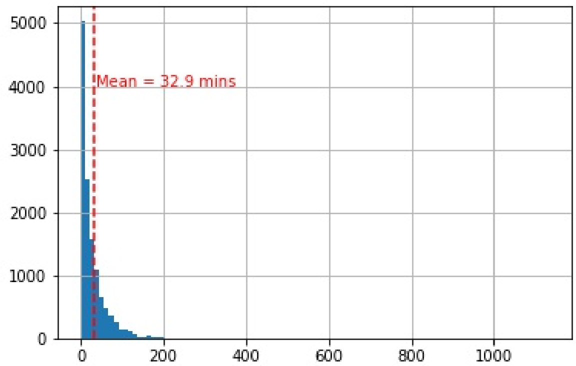


Note: News sources are extracted from each news body, provided by Bloomberg

As for the frequencies of news articles, Figure 3.4 presents that, on average, around every thirty minutes, there is one news article about WTI crude released. The news frequency distribution gives guidance on what should be price frequency to analyze the news effect. Attempt to uncover the news effect, this paper assumes the market is efficient, which means the market reacts fast enough to new information. This paper thus defines any news article's effect on WTI crude oil future price is reflected by how the price was changed within five minutes after this news has been released.

To analyze how the price was changed, this paper establishes a dummy variable based

Figure 3.4: News Frequency Distribution



Note: This graph shows the distribution of how frequently a news has been released in the dataset. The x-axis is in the unit of minute, and the y-axis is the count. On average, around every 33 minutes, there release one news article about WTI crude oil in 2019.

on WTI crude oil future price, indicating a price increase or decrease episode. If price increases within five minutes after a news article has been released, the price dummy will be one. Otherwise, if price decreases or doesn't change, the price dummy will be zero. There is nearly no case for the price to remain the same within the five-minute slot throughout the dataset. Thus, when price dummy equals zero, it means the price has decreased within five minutes after this news release. For each news article, by looking for the price change within five minutes after release, this paper matches news and the price dummy. Figure 3.5 examines whether the data is balanced by comparing the number of price increase episodes with price decrease episodes.

Roughly speaking, the data is balanced, with increasing price episodes slightly larger.

Figure 3.5: Price Dummy Distribution

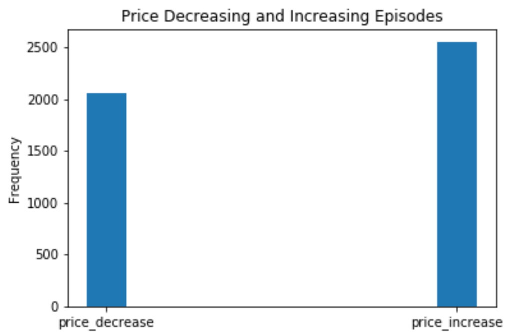
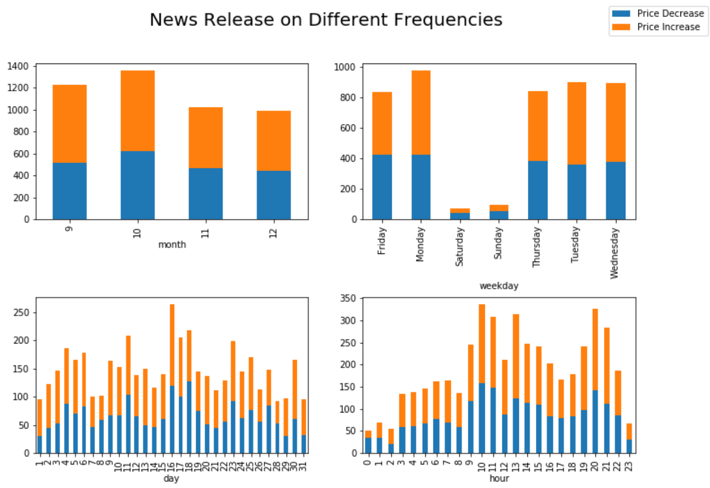
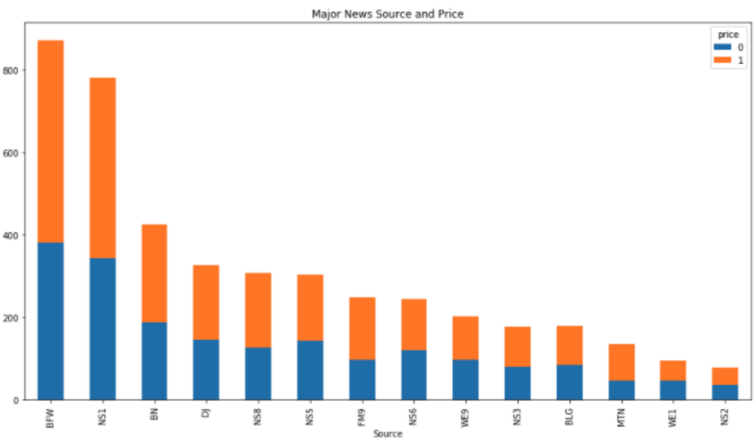


Figure 3.6: Matching News and Price



Except examining the balance of episodes across all datasets, this paper also presents the price sensitivity in different release times of the news articles. In other words, whether news released at certain times, like in the morning or at the end of the month are more likely to have a biased impact on price. Figure 3.6 shows the price decrease and increase episodes distributions across different months, days of the week, date of the month, and hours of the day. The figure indicates that over one thousand news on the topic of crude oil news were released every month. Moreover, most of the news are released on weekdays rather than on the weekends, and most of them are released during the daytime rather than night. However, there are no obvious patterns for the date of the news release. Comparing the number of news following price increase and price decrease episodes, Figure 3.6 presents that there are no significant differences in these two episodes at any time frequencies. Similarly, Figure 3.7 attempts to uncover the relationship between price and news sources. By examining the major news sources, it is clear that all these new sources have almost even numbers of the price increase and decrease episodes across the dataset. Thus, it is essential to analyze the news contents, rather than the time of the news release, or the news sources, to understand the positive or negative news effect on crude oil futures prices.

Figure 3.7: Matching News and Price



3.3.2 Analyze Text Data

In order to analyze the news contents, this paper preprocesses news before further analysis with regressions. Text data are unique data types that need transformation before fitting into a regression model. This paper follows the standard text mining procedures to extract the useful features from the news contents, including tokenization, removing stopwords and lemmatization.

The First step of preprocessing text data is to break every sentence into individual words, which is called tokenization. Taking individual words rather than sentences breaks down the connections between words. However, it is a common method to use to analyze large sets of text data. It is efficient and convenient for computers to analyze the text data by examines what words appear in an article and how many times these words appear, and this analysis is sufficient enough to give insightful results.

After tokenization, each news article will transform into a list of words, symbols, digits, and punctuation. The next step is to remove useless information. For this analysis, symbols, digits, and punctuation are not very useful, so that this paper removes them. Furthermore, this paper removes stopwords. Stopwords are words that frequently appear in many articles, but without significant meanings. Examples of stopwords are 'I', 'the', 'a', 'of'. These are the words will not intervene in the understanding of articles if removed. Besides using the standard English stopwords provided by the NLTK library[2], this paper also includes other lists of stopwords, provided by Loughran and McDonald(2016). These lists of stopwords are widely used in economic analysis, including dates and time, more general words that are not economic meaningful[3].

Removing stopwords, along with symbols, digits, and punctuation, each news article will transform into a list of meaningful words. However, in order to count the appearance of each word, it is essential to remove grammar tense and transform each word into its original form. For example, if we want to calculate how many times the word 'open' appears in a news

[2]NLTK is a python package for text analysis. It contains a list of English stopwords.
[3]Loughran and McDonald's list can be found at https://sraf.nd.edu/textual-analysis/resources/StopWords.

article, we need to count the appearances of 'open', 'opens', 'opened'. Thus, lemmatization is an essential step for text transformation. Lemmatization is taking each word into its original lemma. Another way of converting words is called stemming, which is taking the linguistic root of a word. The reason why this paper chooses lemmatization over stemming is that after stemming, some words become hard to read. For interpretation purposes, the lemma is better than the linguistic root. After lemmatization, each news article will transform into a list of words that are all in their original forms.

Figure 3.8: Original News Article

```
'     Total Refinery, Alpiq, Trade Woes: European Energy Pre-Market  2019-08-26 06:12:07.16 GMT      By John Viljoen
(Bloomberg) -- The following may affect European energy  shares today:  * Note, U.K. markets closed due to holiday
News  * Watch These European Stocks as U.S.-China Trade Woes Escalate  * ALPH SW: Alpiq First Half Adjusted Ebitda CH
F55 Mln  * FP FP: Total Reduces Some Gonfreville Refinery Unit Rates Amid  Strike  * LSNG RM: Lenenergo Second Quarte
r Net Income 3.91 Bln Rubles,  +36% Y/y  * ORSTED DC: Ørsted, Eversource Submit Massachusetts Wind Farm  Proposal
Commodities  * WTI Crude: -1% to $53.64/bbl  * Brent Crude: -0.9% to $58.83/bbl  * Natgas: +1.1% to $2.175/Mmbtu    A
genda  * N.A. Energy Weekly Agenda  * Oil daybook Europe  * Earnings:  ** Other  *** Alpiq Holding AG (ALPH SW)  ***
Maha Energy AB (MAHAA SS)    For more energy wraps in Europe, click here.  For more energy sector wraps in the U.S.,
click here.    To contact the reporter on this story:  John Viljoen in Cape Town at jviljoen@bloomberg.net  To contac
t the editor responsible for this story:  Blaise Robinson at brobinson58@bloomberg.net    '
```

Figure 3.9: News Article after Text Preprocessing

```
'total refinery alpiq trade woe european energy pre market john viljoen bloomberg follow affect european energy share
today note market close due holiday news watch european stocks hina trade woes escalate alph alpiq half adjust ebitda
mln total reduces gonfreville refinery unit rates amid strike lsng lenenergo net bln rubles orste rste eversource sub
mit massachusetts wind farm proposal commodity crude brent crude natgas mmbtu agenda energy agenda oil daybook europe
earning alpiq holding maha energy energy wrap europe click energy sector wrap click contact reporter story john viljo
en cape town contact responsible story blaise robinson'
```

Figure 3.8 and 3.9 shows an example of news article before and after text preprocessing. After tokenization, removing unnecessary words and lemmatization, the original news articles only contains informative words that are ready for further transformation, which will be discussed in Section 3.4.

3.4 News Sentiment Analysis

This section will discuss the news sentiment analysis methodology and how the positive score is calculated for every news article. First, this paper analyze the sentiment of each unique word using a logistic regression. The estimated logistic regression coefficient for each unique word represents its sentiment. This paper defines the effect's direction of each unique word by the sign of its coefficient, and the size of the effect by the absolute value of its coefficient. Moreover, by selecting the words with the highest absolute value in coefficients, this paper defines the most positive and negative words indicating a price increase or decrease episode. Lastly, this paper calculates the positive score for each news article based on how many positive and negative words this news article contains.

3.4.1 News Sentiment Analysis on Unique Words

News sentiment analysis is the analysis that uncovers the predicting power of each unique word in indicating a price increase or decrease episode. In order to do so, this paper uses a supervised machine learning algorithm, which is called logistic regression. Logistic regression is a classification algorithm that deals with binary classification problems. Binary classification has exactly two classes to choose between. In this paper, there are positive and negative classes indicating price increase or price decrease. Logistic regression is a linear classifier, it is a transformation from a linear function:

$$f(x) = b_0 + b_1 * x_1 + ... + b_n * x_n \tag{3.1}$$

where b_0, $b_1 ... b_n$ are the estimators of the regression coefficients for a set of independent variable $x = (x_1, x_2 ... x_n)$. The logistic regression function $p(x)$ is the sigmoid function of $f(x)$:

$$p(x) = \sigma(f(x)) = \frac{1}{1 + exp^{-f(x)}} \tag{3.2}$$

After transformation, $p(x)$ will be in the range of $[0, 1]$, which can be interpreted as probability. Generally, $p(x)$ is interpreted as the predicted probability that $f(x)$ given x is equal to one, and

$1 - p(x)$ is the probability that $f(x)$ is zero. In this paper, $p(x)$ is defined as the probability that WTI crude oil futures price increases within five minutes after news article x_i's release.

Applying logistic regression to conduct news sentiment analysis, this paper treats each news article as a observation, and the contents in news article as the features, and estimates β_{w0}, β_{w1}, ... β_{wj} from the following equation:

$$\begin{pmatrix} Y_0 \\ Y_1 \\ Y_2 \\ ... \\ Y_i \end{pmatrix} = \begin{pmatrix} X_{0,w0} & X_{0,w1} & ... & X_{0,wj} \\ X_{1,w0} & X_{1,w1} & ... & X_{1,wj} \\ X_{2,w0} & X_{2,w1} & ... & X_{2,wj} \\ ... & & & \\ X_{i,w0} & X_{i,w1} & ... & X_{i,wj} \end{pmatrix} * \begin{pmatrix} \beta_{w0} \\ \beta_{w1} \\ ... \\ \beta_{w_j} \end{pmatrix} \tag{3.3}$$

where i stands for each news article as a new observation, and wj is the jth unique word in all news articles. On the left hand side, Y_i is the price change dummy described in the previous section. Specifically, the value of Y is decided by the following conditions:

$$Y_i = \begin{cases} 1, & if\ price_{t+5} - price_t > 0 \\ 0, & otherwise \end{cases} \tag{3.4}$$

On the right hand side, the first term is a sparse matrix, with each row stands for each news article and each column stands for each unique word. There are over 20,606 unique words that has ever shown in 4616 news articles, which indicates the shape of the sparse matrix. Each value $X_{i,wj}$ of the sparse matrix is denoted as the tfidf value for each unique word wj in each news article i. Tfidf is short for term frequency, inverse document frequency. It is a common feature engineering method for text analysis and is widely used in literature. For example Bi and Traum(2020), Fraiberger(2019), Shapiro(2018) have used tfidf to extract text features for different analysis. Specifically, tfidf is calculated by:

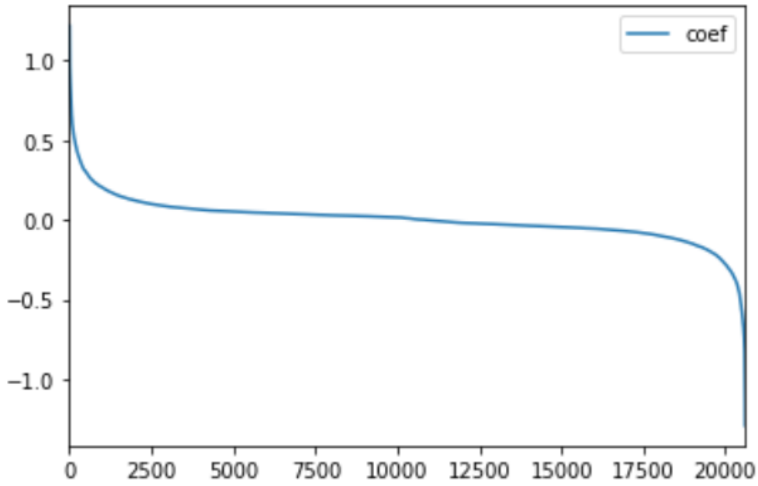
$$X_{i,wj} = \frac{1 + log(t_{i,wj})}{1 + log(\sum_i^N t_{i,wj})} * log(\frac{N}{\sum_{wj} t_{i,wj}}) \tag{3.5}$$

where $t_{i,wj}$ is the frequency of word wj appears in news article i. By examining the equation, it is clear that the first term is the calculating the term frequency and the second term is calculating the inverse document frequency. The first term is evaluating how many time the word wj appear in news article i, normalized by the length of news article i. The higher term frequency indicating a higher tfidf value, presenting the fact that the word wj plays a very important role in news article i by appearing significant times. However, the effect of wj will be weaken if wj also appears in many other news articles besides i, which means it is a common word for this topic. This process is captured by the second term, which is the inverse of how many news articles wj appears divided by the total number of news articles. In this case, N equals to 4616. Combining two effects, a word wj with high tfidf values in news article i means that wj appears many times in news article i, and only appears in few other news articles.

After transforming all preprocessed news articles into the sparse matrix, all data are ready for regression. Figure 3.10 shows the estimation results from Equation 3.3 fitting a logistic regression model. Each point in the x-axis stands for a unique word collecting from all news articles, and there are 20,606 of them. The y-axis stands for the sign and the size of the coefficient for each word. Figure 3.10 indicates that most of the unique word by itself has very limited effect in affecting price, with coefficient very close to zero. However, there are some words that have coefficients with absolute value over 0.5, which are defined as the most positive and negative words by this paper.

For the most positive and most negative words, based on the logistic regression results, 152 words have coefficients smaller than -0.5, and they are collected as the most negative words. On the other side, 159 words have coefficient over 0.5, and are defined as the most positive words. In total, positive and negative words account for around 1.5% of all unique words. Figure 3.11 shows the word clouds of the most positive and most negative words. The font size of each word

Figure 3.10: The Effect of Words



Note: Each point in the x axis stands for a unique word, collecting from all news articles

in the word clouds indicates the size of its coefficient in absolute value. Figure 3.11 shows that words like 'analyst','investing' are showing the greatest effect in indicating a crude oil futures price increase episode, and words like 'partner','hike' contribute the most to crude oil futures price decrease in a news article.

3.4.2 Positive Score for Each News Article

After defining the most positive and negative words in terms of contributing to price increase or decrease episodes, this section will discuss how to calculate the sentiment score for each news article. This paper defines the sentiment score of a news article based on how many positive and negative word it contains. Specifically, this paper calculates the positive score for each news article by comparing how many times positive words appear with the appearance of

Figure 3.11: The Word Clouds

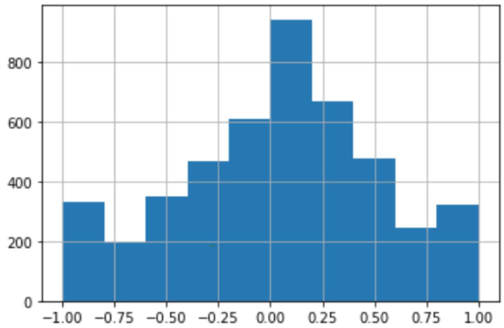
the negative words. Thus, the positive score of a news article i is defines as follow:

$$Score_i = \frac{N_{pos} * F_{pos} - N_{neg} * F_{neg}}{N_{pos} * F_{pos} + N_{neg} * F_{neg}} \tag{3.6}$$

where N_{pos} is the total counts of positive words in news article i, and F_{pos} is the total frequency of each positive/negative word in news article i. For example, if news article i has two positive words, and one appears 10 times and another one appears 15 times, N_{pos} will be two and F_{pos} will be 25. To calculate the positive score, this paper finds the total occurance of positive words and negative words in each news article, takes the difference, and normalizes it by their summation. Figure 3.12 shows the positive score distribution across all news articles.

The positive score of a news article takes a value between $[-1, 1]$. A positive value means this news article is positive, while a negative value means this article is negative. As the absolute value of positive score increases, it becomes more positive or more negative depending on the sign. Figure 3.12 shows that most of the news article has a positive score close to zero, indicating the distribution is roughly a normal distribution with mean at zero, if ignore the two ends. If the positive score for a news article is one, it only has positive words and vice versa.

Figure 3.12: The Positive Score Distribution



Among all 4616 news articles, only around 600 articles are in such cases. Thus, this paper still assume the monotonic positive correlation between the positiveness of a news article and its positive score.

In summary, this section uncovers the effect of each unique word collection from all news articles in the last quarter of 2019, attempts to estimate the coefficients for each unique word on how it affects the crude oil futures prices. In order to do so, this paper uses a supervised machine learning algorithm called logistic regression to solve this binary classification problem. Logistic regression estimates the coefficients for all unique words and selects the most positive and negative words based on the coefficients' sign and size. The most positive and negative words help identify the sentiment score for each news article. By calculating the total occurrence of positive and negative words in a news article, this paper defines the positive score for each news article, which is useful for the regression in Section 3.6.

3.5 Topic Analysis

Instead of reading each article and manually separating our news sample into different topics, we use an unsupervised machine learning algorithm, K-means, to detect common patterns in news articles and group them into clusters, i.e., topics.

3.5.1 K-means Algorithm

K-means is one of the most commonly used clustering algorithms in machine learning. Unlike supervised learning, which first defines a list of keywords in each topic and classify each article that includes those keywords into a specific topic, researchers first need to determine the number of clusters (topics), then the K-means algorithm will assign each sample to the cluster where its distance from the centroid of the cluster is minimized. The algorithm of K-means is described as followed:

1. Specify the number of K clusters. [4]

2. Initialize the centroid point $\mu_k (k \in K)$ of each cluster with a random value.

3. Calculate the squared Euclidean distance of each sample x_i to the centroid point of each cluster.

$$||x_i - \mu_k||^2$$

4. Assign the sample x_i to the closest cluster k where the squared Euclidean distance is minimized.

$$\min_k ||x_i - \mu_k||^2$$

5. Update μ_k by taking the mean of sample points assigned to cluster k.

[4]In theory, the Elbow method helps to determine the optimal K values, by plotting K on the horizontal axis and sum of minimized distances of each cluster on the vertical axis. The optimal K is found when the slope of Elbow curve flattens, i.e., when the y value converges. However, in practice, the Elbow curve does not converge so that users need to determine the value of K based on their purpose. Our result of Elbow method can be provided upon request.

6. Repeat 3-5 until the sum of the squared distances overall K clusters is minimized.

$$\min \sum_{k=1}^{K} \sum_{x_i \in c_k} ||x_i - \mu_k||^2$$

where c_k denotes sample points in cluster k.

Following Bi and Traum(2019), we use the tfidf vector of each article (i.e., row vector of the tfidf matrix discussed in the previous section) as a sample point for training the algorithm, so that each article will be assigned uniquely to a cluster. The training set covers the whole sample period (20190101 - 20191231), which includes a total of 13,183 news.

When determining the value of K, the Elbow method doesn't provide us an optimal choice of K (the Elbow curve doesn't converge). Therefore, we run the algorithm several times with different K values and thus choose $K = 4$ based on the results that provide us with the most meaningful and best interpretability of news topics.

3.5.2 Clustering Results

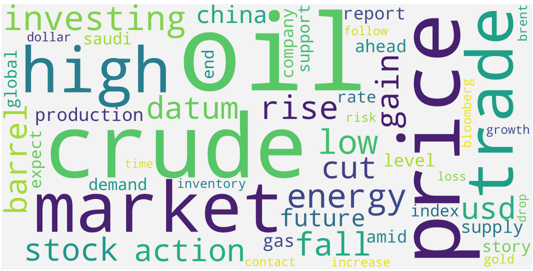
The K-means algorithm generates 4 clusters of news topics. Figure 3.13 plots the word clouds of top 50 important words[5] for each topic, with the more important ones shown in a bigger font. Given our specification of four news topics, we are able to interpret each topic to be: (a) - "World Crude Oil", (b) - "WTI Crude Oil", (c) - "Financial Analysis" and (d) - "Editorial Opinion". The "World Crude Oil" cluster has keywords related to the global oil market, such as market, trade, energy, China, USD, etc. "WTI Crude Oil" cluster includes keywords like price, supply, demand, production, which reflects more about the information of WTI crude oil in the North American region[6]. "Financial Analysis" contains keywords position, spread, call,

[5]For each cluster, we sum each word's tfidf value over all the articles within this cluster, rank them from the highest to the lowest, and then pick the top 50 words for word cloud plot.

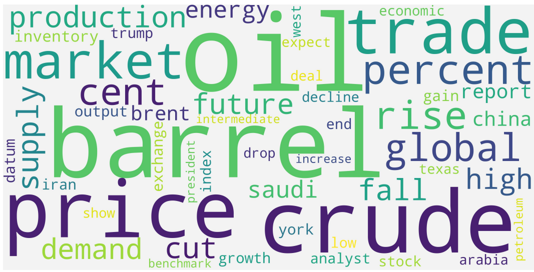
[6]There are some words shown in both of these two topics, which interprets these two topics to be unclear and need to be improved in future work.

put, option, relating to the price analysis, and use of derivatives for crude oil. The last cluster is "Editorial Opinion," with keywords story, contact, reporter, etc.

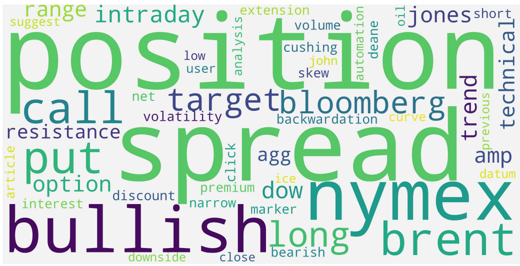
Figure 3.13: Word clouds of Clustering Results



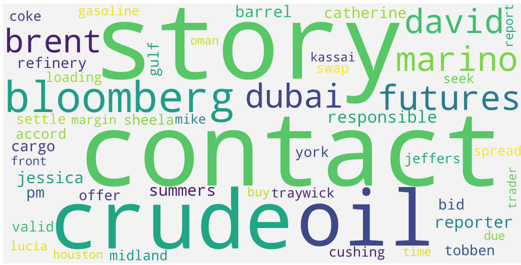
(a) World Crude Oil



(b) WTI Crude Oil



(c) Financial Analysis



(d) Editorial Opinion

Table 3.1 shows the number of news in each topic. More than 80% of news throughout year 2019 are assigned to the “World Crude Oil” and “WTI Crude Oil” topics. 13.86% news are related to “Editorial Opinion”, which is more likely to be a summary or report of already released new information. There are only 563 news classified into “Financial Analysis” topic, which includes both analysis of price change and analysts’ forecast of future price movement. As we can see in the next section, the last two topics are less important indicators of price changes.

Table 3.1: Number of News in Each Topic

Topic Index	**Topic Name**	**news_cnt**	**%**
a	World Crude Oil	6,386	48.44%
b	WTI Crude Oil	4,407	33.43%
c	Financial Analysis	563	4.27%
d	Editorial Opinion	1,827	13.86%
total		13,183	100%

3.6 Test News Topic and Sentiment Score on Futures Price Change

In this section, we estimate the relation of news topic and sentiment score with futures price change using logistic regression. By assuming that the WTI crude oil futures market is nearly perfectly efficient, investors responding quickly to any new information released to the market, futures price should adjust quickly in line with its intrinsic value. For example, a good news article about the discovery of new oil wells should push down the futures price immediately, as investors are competing with each other to find this price decrease opportunity. Since the market is nearly efficient, the futures price can reflect this new information very quickly. Given this assumption, we can test the relation between crude oil news released and its futures price change.

We first define our dependent variable to be one if futures price after 5 minutes is higher than the price when news released; and zero otherwise[7], as stated in Section 3.3. In future work, we will use different time spam, calculating price change for robustness checks. To test how crude oil news relates to price changes, we use the logistic regression, which estimates the influence of x on Y's probability for each news article. The regression equation we estimate

[7] The price data we use is high-frequency data, with 5 minutes interval. In our final regression sample (2019Q4), there doesn't exist price unchanged within each 5 minutes interval.

follows:

$$Prob(Y_i = 1|x) = \sigma(\beta_0 + \beta_1 Score_i + \sum_{k}^{3} \beta_k Topic_k + \sum_{k}^{3} \beta_{k,i} Topic_k * Score_i) \tag{3.7}$$

where

$$Y_i = \begin{cases} 1, & if \quad price_{t+5} > price_t \\ 0, & otherwise \end{cases} \tag{3.8}$$

and $Topic$, a categorical variable, is {0: "World Crude Oil"; 1: "WTI Crude Oil"; 2: "Financial Analysis"; 4: "Editorial Opinion"}. In the regression, three dummy variables are generated indicating which topic this news article is in, and topic 0('World Crude Oil') is chosen as the benchmark topic. As described in section 3.4, $score$ evaluates the degree of news sentiment. The higher the $Score$ is, the more positive the news is.

The LHS of equation (3.7) denotes the probability of price increase ($Y = 1$) given all x. As discussed in Section 3.4.1, on the RHS, $\sigma(\cdot)$ is a Sigmoid function. In the parenthesis of $\sigma(\cdot)$ is the linear combination of independent variables that we are interested.

Our final logistic regression sample covers only the last quarter of 2019, with 4,616 observations due to the availability of high-frequency futures price data. Table 3.2 reports the regression results. The coefficient on constant is interpreted as the impact of news in topic 0 ("World Crude Oil") on the probability of price increase. On average, news in the "World Crude Oil" topic is more likely to be correlated with price increases. If the news is related to "WTI Crude Oil," the coefficient becomes 0.0159 (0.3267 minuses 0.3108), meaning that "WTI Crude Oil" news on average doesn't have a significant impact on a price increase. For topic 2 & 3 ("Financial Analysis" and "Editorial Opinion"), news has relatively less positive impacts on price increases, as their coefficients (-0.2821 and -0.2581) are both statistically significant and smaller than the coefficient of topic 0 (0.3267).

After considering the impact of sentiment score for each news under different topics, the coefficient on $Score$ can be interpreted as a correlation between the degree of topic 0 ("World

Crude Oil") news sentiment and the probability of price increase. The result is statistically significantly positive (1.2642), meaning that the more positive "World Crude Oil" news is, the higher probability that price will increase within 5 minutes. The coefficient on the interaction term $Topic1 * Score$ (0.6034), is also positive and statistically significant, which implies that the sentiment score of news under topic 1 ("WTI Crude Oil") is more correlated with the probability of price increase than news under topic 0 ("World Crude Oil"). Coefficients of the interaction terms $Topic2 * Score$ and $Topic3 * Score$ are negative but not statistically significant. Thus we reject the hypothesis that news sentiment under topic 2 and 3 is statistically different from those under topic 0.

To sum up, on average, "World Crude Oil" news has the highest correlation with a price increase. Nonetheless, the more positive news is under the topic "WTI Crude Oil," the higher probability that WTI crude oil futures price will increase within 5 minutes.

3.7 Conclusion

With the development of machine learning algorithm and its use in economics literature, it is worthwhile to apply machine learning algorithm to understand news impact on financial asset price movement. In this project, we use both supervised and unsupervised machine learning algorithms to learn impact of news sentiment and news topics on crude oil futures price increase. By assuming that crude oil futures market are nearly perfectly efficient (price adjusts quickly to any new information released to the public), we use high frequency data for estimating the news impact. The results show that "World Crude Oil" news on average is positively correlated with price increase, and the more positive "WTI Crude Oil" news is, the higher probability that crude oil futures price will increase within five minutes. The implication of our project is that

Table 3.2: Test for News Effect on Price Change

$$Prob(Y = 1|X) = \sigma(\beta_0 + \beta_1 Score + \beta_2 Topic + \beta_3 Topic * Score)$$

Dependent Variable	**Coef.**	
2[1]*Score	1.2642	***
	(0.082)	
2[0]*Topic1 "WTI Crude Oil"	-0.3108	***
	(0.073)	
2[0]*Topic2 "Financial Analysis"	-0.2821	*
	(0.161)	
2[0]*Topic3 "Editorial Opinion"	-0.2581	**
	(0.105)	
2[0]*Topic1 * Score	0.6034	***
	(0.201)	
2[0]*Topic2 * Score	-0.1910	
	(0.429)	
2[0]*Topic3 * Score	-0.0868	
	(0.205)	
2[1]*Const.	0.3267	***
	(0.044)	
No. Observations	4,616	
Pseudo R-squ.	0.071	
Robust Std. Err.	YES	

*Notes: The number in the parenthesis reports the t-statistics. *** indicates* $P < 0.01$. *** indicates* $P < 0.05$. ** indicates* $P < 0.1$.

using the coefficients in our last regression results, we are able to construct a news index, which can be used further in estimating the magnitude of price change, together with other macro and micro economic control variables.

There are a lot of work can be done in the future to improve our results. For example, we will modify our filter of stop word in order to improve our sentiment results to be better consistent with human cognition. We can also try Neural Network algorithm to learn news topic, which as a supervised machine learning algorithm might provide more precise learning than the K-means algorithm that we used in the project.

Bibliography

[1] Joseph V. Balagtas and Matthew T. Holt. The Commodity Terms of Trade, Unit Roots, and Nonlinear Alternatives: A Smooth Transition Approach. *American Journal of Agricultural Economics*, 91(1):87–105, 2009.

[2] Brad M Barber and Terrance Odean. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The review of financial studies*, 21(2):785–818, 2008.

[3] Christiane Baumeister and Lutz Kilian. A general approach to recovering market expectations from futures prices with an application to crude oil. 2014.

[4] L. Bettendorf and H Dewachter. *Ageing and the Relative Price of Nontrad- ables.* Tinbergen Institute Discussion Paper.

[5] Huixin Bi and Nora Traum. Sovereign risk and fiscal information: A look at the us state default of the 1840s. *Federal Reserve Bank of Kansas City Working Paper*, (19-04), 2019.

[6] Michael F Bleaney and David Greenaway. Long-Run Trends in the Relative Price of Primary Commodities and in the Terms of Trade of Developing Countries. *Oxford Economic Papers*, 45(3):349–63, 1993.

[7] Michael W Brandt and Lin Gao. Macro fundamentals or geopolitical events? a textual analysis of news events for crude oil. *Journal of Empirical Finance*, 51:64–94, 2019.

[8] Joseph P. Byrne, Giorgio Fazio, and Norbert Fiess. Primary commodity prices: Co-movements, common factors and fundamentals. *Journal of Development Economics*, 101(C):16–26, 2013.

[9] Joseph P Byrne, Giorgio Fazio, and Norbert Fiess. Primary commodity prices: Co-movements, common factors and fundamentals. *Journal of Development Economics*, 101:16–26, 2013.

[10] Paul Cashin, Hong Liang, and C. John McDermott. How Persistent Are Shocks to World Commodity Prices? *IMF Staff Papers*, 47(2):2, 2000.

[11] Paul Cashin, Hong Liang, and C John McDermott. How persistent are shocks to world commodity prices? *IMF Staff Papers*, 47(2):177–217, 2000.

[12] Paul Cashin and C. John McCDermott. The Long-Run Behavior of Commodity Prices: Small Trends and Big Variability. *IMF Staff Papers*, 49(2):2, 2002.

[13] Paul Cashin, C. John McDermott, and Catherine Pattillo. Terms of trade shocks in Africa: are they short-lived or long-lived? *Journal of Development Economics*, 73(2):727–744, 2004.

[14] Paul Cashin, C. John McDermott, and Alasdair Scott. Booms and slumps in world commodity prices. *Journal of Development Economics*, 69(1):277–296, 2002.

[15] Paul Cashin, C John McDermott, and Alasdair Scott. Booms and slumps in world commodity prices. *Journal of development Economics*, 69(1):277–296, 2002.

[16] Nelson C. Mark Cecchetti, Stephen G. and Robert J. Sonora. Price index convergence among united states cities. *International Economic Review*, 43(4):1081–1099, 2002.

[17] Wesley S Chan. Stock price reaction to news and no-news: drift and reversal after headlines. *Journal of Financial Economics*, 70(2):223–260, 2003.

[18] Shu-Ling Chen, John D. Jackson, Hyeongwoo Kim, and Pramesti Resiandini. What drives commodity prices? *American Journal of Agricultural Economics*, 96(5):1455–1468, 2014.

[19] Yu-Chin Chen, Kenneth Rogoff, and Barbara Rossi. Can exchange rates forecast commodity prices? *The Quarterly Journal of Economics*, 125(3):1145–1194, 2010.

[20] Yung-Ping Chen and Kwang-Wen Chu. Household expenditure patterns: the effect of age of family head. *Journal of Family Issue*, pages 233–250, 1982.

[21] Yin-Wong Cheung, Kon S. Lai, and Michael Bergman. Dissecting the ppp puzzle: the unconventional roles of nominal exchange rate and price adjustments. *Journal of International Economics*, 64(1):135–150, 2004.

[22] Lawrence J. Christiano, Martin Eichenbaum, and Charles L. Evans. Monetary policy shocks: What have we learned and to what end? In J. B. Taylor and M. Woodford, editors, *Handbook of Macroeconomics*, volume 1 of *Handbook of Macroeconomics*, chapter 2, pages 65–148. Elsevier, 1999.

[23] J. Coto-Martinez and J. C Reboredo. The relative price of non-traded goods under imperfect competition. *Oxford Bulletin of Economics and Statistics*, 76:24–40, 2014.

[24] John T. Cuddington. Long-run trends in 26 primary commodity prices : A disaggregated look at the Prebisch-Singer hypothesis. *Journal of Development Economics*, 39(2):207–227, 1992.

[25] Giovannini A. andWolf H. C De Gregorio, J. International evidence on tradables and nontradables inflation. *European Economic Review*, 38:1225–1244, 1994.

[26] Angus Deaton. Commodity Prices and Growth in Africa. *Journal of Economic Perspectives*, 13(3):23–40, 1999.

[27] Angus Deaton and Guy Laroque. On the Behaviour of Commodity Prices. *Review of Economic Studies*, 59(1):1–23, 1992.

[28] Angus Deaton and Guy Laroque. On the behaviour of commodity prices. *The review of economic studies*, 59(1):1–23, 1992.

[29] John J. Seater. DeJuan, Joseph P. The permanent income hypothesis: Evidence from the consumer expenditure survey. *Journal of Monetary Eco- nomics*, 43(2):351–376, May 1999.

[30] Elizabeth Demers, Clara Vega, et al. Soft information in earnings announcements: News or noise? 2008.

[31] John Elder, Hong Miao, and Sanjay Ramchander. Jumps in oil prices: the role of economic news. *The Energy Journal*, 34(3), 2013.

[32] Charles Engel and James Morley. The adjustment of prices and the adjustment of the exchange rate. NBER Working Papers 8550, National Bureau of Economic Research, Inc, 2001.

[33] R. C. Fair and K. M. Dominguez. Effects of the changing u.s. age distribution on macroeconomic equations. *American Economic Review*, 81:1276–1294, 1991.

[34] Jeffrey A. Frankel. The Effect of Monetary Policy on Real Commodity Prices. In *Asset Prices and Monetary Policy*, NBER Chapters, pages 291–333. National Bureau of Economic Research, 2008.

[35] Diego Garcia. Sentiment during recessions. *The Journal of Finance*, 68(3):1267–1300, 2013.

[36] Atanu Ghoshray. A Reexamination of Trends in Primary Commodity Prices. *Journal of Development Economics*, 95(2):242–251, 2011.

[37] Atanu Ghoshray. Dynamic Persistence of Primary Commodity Prices. *American Journal of Agricultural Economics*, 95(1):153–164, 2013.

[38] Enzo R Grilli and Maw Cheng Yang. Primary Commodity Prices, Manufactured Goods Prices, and the Terms of Trade of Developing Countries: What the Long Run Shows. *World Bank Economic Review*, 2(1):1–47, 1988.

[39] Christoph Kaufmann. Groneck Max. Determinants of relative sectoral prices: The role of demographic change. *Oxford Bulletin of Economics and Statistics*, 79(3):319–347, 2017.

[40] James D Hamilton. Understanding crude oil prices. *The Energy Journal*, 30(2), 2009.

[41] J. Michael Harris and Noel Blisard. Food spending and the elderly. *Food Review*, pages 14–18, 2002.

[42] David I. Harvey, Neil M. Kellard, Jakob B. Madsen, and Mark E. Wohar. The Prebisch-Singer Hypothesis: Four Centuries of Evidence. *The Review of Economics and Statistics*, 92(2):367–377, 2010.

[43] Rodolfo Helg. A note on the stationarity of the primary commodities relative price index. *Economics Letters*, 36(1):55–60, 1991.

[44] Gong-Soog Hong and SooYeon Kim. Out-of-pocket health care expenditure patterns and financial burden across the life cycle stages. *Journal of Consumer Affairs*, pages 291–313, 2000.

[45] Chiu Fui Joyce Mok Jinkook Lee, Sherman D. Hanna and HuiWang. Apparel expenditure patterns of elderly consumers: a life-cycle consumption model. *Family and Consumer Sciences Research Journal*, 26(2):109–140, 1997.

[46] Neil Kellard and Mark E. Wohar. On the Prevalence of Trends in Primary Commodity Prices. *Journal of Development Economics*, 79(1):146–167, 2006.

[47] Rose M. Rubin Kenneth Koelln and Marion Smith Picard. Vulnerable older households: expenditures on necessities by older americans. *Social Science Quarterly*, pages 619–633, 1995.

[48] Lutz Kilian. Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review*, 99(3):1053–69, 2009.

[49] Lutz Kilian and Clara Vega. Do energy prices respond to us macroeconomic news? a test of the hypothesis of predetermined energy prices. *Review of Economics and Statistics*, 93(2):660–671, 2011.

[50] Sang-Hee Sohn Eunyoung Rhee Yoon G. Lee Lee, Seonglim and Hua Zan. Consumption patterns and economic status of older households in the united states. Technical Report September, Bureau of Labor Statistics, Monthly Labor Review, 2014.

[51] Sherman D. Hanna Chiu Fui Joyce Mok Lee, Jinkook and Hui Wang. Apparel expenditure patterns of elderly consumers: A life-cycle consumption model. *Family and Consumer Sciences Research Journal*, 26(2):109–140, 1997.

[52] Tim Loughran and Bill McDonald. Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4):1187–1230, 2016.

[53] M. Luhrmann. *Population Aging and the Demand for Goods and Services.* MEA Discussion Paper.

[54] Paul Newbold, Stephan Pfaffenzeller, and Anthony Rayner. How well are long-run commodity price series characterized by trend components? *Journal of International Development*, 17(4):479–494, 2005.

[55] Ral Prebisch. The Economic Development of Latin America and Its Principal Problems. Technical report, Economic Commission for Latin America, Department of Economic Affairs, United Nations, April 1950.

[56] Avihai Rapaport. Supply and demand shocks in the oil market and their predictive power. *Available at SSRN 2472379*, 2010.

[57] Kenneth Rogoff. The purchasing power parity puzzle. *Journal of Economic Literature*, 34(2):647–668, 1996.

[58] D Sapsford. The Statistical Debate on the Net Barter Terms of Trade between Primary Commodities and Manufactures: A Comment and Some Additional Evidence. *Economic Journal*, 95(379):781–88, 1985.

[59] Adam Hale Shapiro, Moritz Sudhof, and Daniel Wilson. Measuring news sentiment. Federal Reserve Bank of San Francisco, 2020.

[60] Robert J Shiller. *Irrational exuberance: Revised and expanded third edition.* Princeton university press, 2015.

[61] Hans W. Singer. The Distribution of Gains Between Investing and Borrowing Countries. *American Economic Review*, 40(2):473–485, 1950.

[62] Cindy Soo. Quantifying animal spirits: news media and sentiment in the housing market. *Ross School of Business Paper*, (1200), 2015.

[63] Paul C Tetlock. Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance*, 62(3):1139–1168, 2007.

[64] Kenneth West and Ka-Fu Wong. A factor model for co-movements of commodity prices. *Journal of International Money and Finance*, 42(C):289–309, 2014.

[65] Jeffrey C. Williams and Brian D. Wright. *Storage and Commodity Markets.* Number 9780521326162 in Cambridge Books. Cambridge University Press, Cambridge, UK, 1991.

[66] Ma. Xiaofei. *Demographic Factor and the Rise of Services.* Working Paper.

Appendix A

Chapter Two

A.0.1 Dynamic elasticity with shocks

Table A.1: IMF Codes of the World Commodity Prices

ID	IMF Code	Commodity	ID	IMF Code	Commodity
1	PBARL	Barley, Spot Price	26	PCOFFORB	Coffee, Robusta, New York Cash Price
2	PGNUTS	Peanuts, CIF Argentina	27	PTEA	Tea, Kenya Auctions, CIF UK
3	PMAIZMT	Maize, FOB Gulf of Mexico	28	PLOGORE	Soft Logs, US Export Price
4	PRICENPQ	Rice, Thailand Price Quote	29	PLOGSK	Hard Logs, Import Price Japan
5	PSMEA	Soybean Meal, Chicago Futures	30	PSAWMAL	Hard Sawnwood, C&F UK Fort
6	PSOYB	Soybeans, Chicago Futures	31	PSAWORE	Soft Sawnwood, US Price
7	PWHEAMT	Wheat, Kansas City	32	PCOTTIND	Cotton, CIF Riverpool
8	PROIL	Rapeseed, FOB Rotterdam	33	PWOOLC	Wool, Coarse, Australian Spot Quotes
9	POLVOIL	Olive Oil, UK Ex-Tanker Price	34	PWOOLF	Wool, Fine, Australian Spot Quotes
10	PPOIL	Palm Oil, Malaysian Futures	35	PRUBB	Rubber, Singapore Comdty Exchange
11	PSOIL	Soybean Oil, Chicago Futures	36	PHIDE	Hides, US FOB Shipping Point
12	PSUNO	Sunflower Oil, US Export Price	37	PALUM	Aluminum, CIF UK Ports
13	PBEEF	Beef, CIF US Import Price	38	PCOPP	Copper, CIF European Ports
14	PLAMB	Lamb, Smithfield London	39	PIORECR	Iron Ore, China CFR Tianjin Port
15	PPORK	Swine (Pork), US Price	40	PLEAD	Lead, CIF European Ports
16	PPOULT	Poultry, US Georgia Docks	41	PNICK	Nickel, CIF European Ports
17	PFISH	Fishmeal, Peru, CIF	42	PTIN	Tin, LME Spot Price
18	PSALM	Salmon, Norwegian Export Price	43	PURAN	Uranium, Nuexco Exchange Spot
19	PSHRI	Shrimp, New York Port	44	PZINC	Zinc, High Grade 98% Pure
20	PBANSOP	Bananas, FOB US Ports	45	PCOALAU	Coal, FOB Newcastle/Port Kembla
21	PORANG	Oranges, CIF French Import Price	46	POILAPSP	Crude Oil, Avg of Brent, Dubai, WTI
22	PSUGAISA	Sugar, CSCE Contract Futures	47	POILBRE	Oil, Brent, FOB UK
23	PSUGAUSA	Sugar, USA Import Price	48	POILDUB	Oil, Dubai, FOB Dubai
24	PCOCO	Cocoa Beans, CIF US & Euro Ports	49	POILWTI	Oil, West Texas Intermediate
25	PCOFFOTM	Coffee, Arabica, NY Cash Price			

Note: All commodity prices are denominated in the US dollar and are obtained from the IMF website. Observations are from 1980:I to 2013:IV. We transformed monthly data to quarterly frequency data by taking end of period values. For more detailed information, refer to the IMF website, https://www.imf.org/external/np/res/commod/.

Table A.2: Contemporaneous Dynamic Elasticity with respect to the Exchange Rate

$$\eta_e^p(0) = \frac{\psi_e^p(0)}{\psi_e^e(0)},\ \psi_e^p(0) = \rho_e^p(0),\ \psi_e^e(0) = \rho_e^e(0)$$

ID	IMF Code	$\eta_e^p(0)$	5%	50%	95%	ID	IMF Code	$\eta_e^p(0)$	5%	50%	95%
1	PBARL	-0.699	-1.297	-0.749	-0.202	26	PCOFFORB	-0.705	-1.296	-0.709	-0.071
2	PGNUTS	-0.658	-1.381	-0.714	-0.021	27	PTEA	-0.834	-1.387	-0.832	-0.317
3	PMAIZMT	-0.210	-0.717	-0.256	0.268	28	PLOGORE	-0.111	-0.429	-0.116	0.183
4	PRICENPQ	-0.116	-0.597	-0.148	0.334	29	PLOGSK	-0.910	-1.349	-0.903	-0.491
5	PSMEA	-0.709	-1.235	-0.724	-0.274	30	PSAWMAL	-0.716	-1.082	-0.697	-0.348
6	PSOYB	-0.637	-1.122	-0.653	-0.215	31	PSAWORE	0.307	0.037	0.303	0.570
7	PWHEAMT	-0.721	-1.245	-0.722	-0.296	32	PCOTTIND	-0.329	-0.764	-0.349	0.108
8	PROIL	-0.928	-1.489	-0.944	-0.393	33	PWOOLC	-0.674	-1.104	-0.692	-0.251
9	POLVOIL	-1.156	-1.481	-1.161	-0.873	34	PWOOLF	-0.518	-0.956	-0.537	-0.099
10	PPOIL	-0.653	-1.304	-0.692	-0.031	35	PRUBB	-1.217	-1.773	-1.206	-0.736
11	PSOIL	-0.453	-1.000	-0.482	-0.006	36	PHIDE	-0.230	-0.809	-0.207	0.314
12	PSUNO	-0.392	-1.061	-0.412	0.186	37	PALUM	-1.226	-1.663	-1.233	-0.719
13	PBEEF	-0.044	-0.298	-0.050	0.169	38	PCOPP	-1.668	-2.286	-1.666	-1.105
14	PLAMB	-0.877	-1.165	-0.886	-0.595	39	PIORECR	0.175	-0.258	0.153	0.539
15	PPORK	0.449	-0.297	0.468	1.184	40	PLEAD	-1.170	-1.867	-1.201	-0.508
16	PPOULT	0.108	-0.073	0.103	0.269	41	PNICK	-1.045	-1.822	-1.057	-0.235
17	PFISH	-0.582	-0.908	-0.586	-0.225	42	PTIN	-0.54	-1.147	-0.573	0.014
18	PSALM	-1.227	-1.592	-1.248	-0.858	43	PURAN	-0.123	-0.567	-0.144	0.380
19	PSHRI	-0.136	-0.451	-0.149	0.146	44	PZINC	-0.753	-1.414	-0.802	-0.181
20	PBANSOP	0.832	-0.203	0.776	1.712	45	PCOALAU	-0.500	-0.900	-0.520	-0.122
21	PORANG	-0.364	-1.345	-0.387	0.579	46	POILAPSP	-1.036	-1.972	-1.055	-0.200
22	PSUGAISA	-1.255	-2.007	-1.277	-0.544	47	POILBRE	-1.165	-2.114	-1.190	-0.336
23	PSUGAUSA	-0.308	-0.554	-0.312	-0.077	48	POILDUB	-0.998	-1.949	-1.016	-0.112
24	PCOCO	-0.709	-1.106	-0.724	-0.301	49	POILWTI	-0.954	-1.906	-0.952	-0.101
25	PCOFFOTM	-0.406	-1.104	-0.398	0.266	Mean: -0.588, Median: -0.658					

Note: p% denotes the p^{th}percentile obtained from 500 nonparametric bootstrap simulations. 50% is the median elasticity estimate. 5% and 95% constitutes the 90% nonparametric confidence band.

Table A.3: Long-Run Dynamic Elasticity with respect to the Exchange Rate

$$\eta_e^p(\infty) = \frac{\psi_e^p(\infty)}{\psi_e^e(\infty)},\ \psi_e^p(\infty) = \sum_{s=0}^{\infty} \rho_e^p(s),\ \psi_e^e(\infty) = \sum_{s=0}^{\infty} \rho_e^e(s)$$

ID	IMF Code	$\eta_e^p(\infty)$	5%	50%	95%	ID	IMF Code	$\eta_e^p(\infty)$	5%	50%	95%
1	PBARL	-1.584	-2.592	-1.569	-0.790	26	PCOFFORB	-1.345	-2.253	-1.385	-0.419
2	PGNUTS	-2.123	-3.407	-2.129	-1.006	27	PTEA	-0.655	-1.643	-0.667	0.175
3	PMAIZMT	-1.231	-2.192	-1.231	-0.368	28	PLOGORE	-0.151	-0.683	-0.135	0.362
4	PRICENPQ	-1.609	-2.434	-1.629	-0.823	29	PLOGSK	-0.912	-1.726	-0.904	-0.158
5	PSMEA	-1.220	-2.053	-1.229	-0.398	30	PSAWMAL	-0.939	-1.611	-0.943	-0.310
6	PSOYB	-1.278	-2.148	-1.306	-0.561	31	PSAWORE	-0.016	-0.326	-0.016	0.336
7	PWHEAMT	-1.078	-1.972	-1.070	-0.382	32	PCOTTIND	-0.790	-1.685	-0.824	0.165
8	PROIL	-1.034	-2.064	-1.076	-0.043	33	PWOOLC	-1.004	-1.671	-1.004	-0.335
9	POLVOIL	-1.125	-1.776	-1.147	-0.478	34	PWOOLF	-1.003	-1.795	-1.024	-0.193
10	PPOIL	-0.667	-1.894	-0.704	0.469	35	PRUBB	-1.610	-2.312	-1.622	-0.961
11	PSOIL	-0.905	-1.877	-0.934	-0.037	36	PHIDE	-0.586	-1.288	-0.551	0.043
12	PSUNO	-1.620	-2.999	-1.695	-0.441	37	PALUM	-1.515	-2.149	-1.518	-0.933
13	PBEEF	-0.205	-0.565	-0.231	0.181	38	PCOPP	-1.562	-2.479	-1.501	-0.665
14	PLAMB	-1.008	-1.563	-1.014	-0.479	39	PIORECR	-0.652	-1.370	-0.660	0.006
15	PPORK	-0.044	-1.003	-0.011	0.925	40	PLEAD	-1.157	-2.330	-1.120	0.119
16	PPOULT	-0.168	-0.432	-0.167	0.071	41	PNICK	-1.570	-2.873	-1.605	-0.416
17	PFISH	-0.475	-1.286	-0.454	0.347	42	PTIN	-0.681	-1.893	-0.723	0.313
18	PSALM	-0.796	-1.271	-0.797	-0.219	43	PURAN	-1.057	-2.140	-1.028	0.111
19	PSHRI	-0.240	-0.850	-0.278	0.345	44	PZINC	-0.577	-1.717	-0.594	0.581
20	PBANSOP	-0.374	-1.172	-0.365	0.381	45	PCOALAU	-1.967	-3.035	-1.949	-0.983
21	PORANG	-0.449	-1.309	-0.462	0.470	46	POILAPSP	-0.909	-2.054	-0.880	0.094
22	PSUGAISA	-1.705	-2.845	-1.672	-0.405	47	POILBRE	-0.994	-2.234	-0.945	0.084
23	PSUGAUSA	-0.440	-0.996	-0.444	0.115	48	POILDUB	-0.864	-2.035	-0.840	0.128
24	PCOCO	0.018	-0.796	-0.018	0.878	49	POILWTI	-0.862	-2.038	-0.873	0.162
25	PCOFFOTM	-1.145	-2.231	-1.194	-0.064		Mean: -0.936, Median: -0.939				

Note: $p\%$ denotes the p^{th}percentile obtained from 500 nonparametric bootstrap simulations. 50% is the median elasticity estimate. 5% and 95% constitutes the 90% nonparametric confidence band. Long-run responses are obtained by taking the 40^{th}period ahead response function estimate.

Table A.4: Contemporaneous Dynamic Elasticity with respect to the Real GDP

$$\eta_y^p(0) = \frac{\psi_y^p(0)}{\psi_y^y(0)},\ \psi_y^p(0) = \rho_y^p(0),\ \psi_y^y(0) = \rho_y^y(0)$$

ID	IMF Code	$\eta_y^p(0)$	5%	50%	95%	ID	IMF Code	$\eta_y^p(0)$	5%	50%	95%
1	PBARL	1.105	-1.149	1.102	3.950	26	PCOFFORB	0.475	-1.689	0.302	2.845
2	PGNUTS	3.346	0.953	3.431	6.058	27	PTEA	-0.643	-3.188	-0.677	1.781
3	PMAIZMT	0.276	-1.465	0.239	2.106	28	PLOGORE	-0.234	-1.815	-0.191	0.996
4	PRICENPQ	-0.113	-3.863	-0.188	2.240	29	PLOGSK	-1.707	-4.216	-1.497	1.067
5	PSMEA	0.195	-1.782	0.251	3.042	30	PSAWMAL	-1.509	-4.219	-1.266	1.582
6	PSOYB	0.952	-0.938	0.973	3.333	31	PSAWORE	-0.018	-1.609	-0.022	1.113
7	PWHEAMT	-0.052	-1.899	-0.160	2.210	32	PCOTTIND	-1.155	-3.169	-1.071	1.523
8	PROIL	3.662	1.229	3.711	6.302	33	PWOOLC	1.350	-0.235	1.424	3.495
9	POLVOIL	-1.635	-3.258	-1.673	-0.342	34	PWOOLF	3.409	1.022	3.309	5.311
10	PPOIL	0.802	-2.180	0.786	5.297	35	PRUBB	2.311	-0.163	2.257	6.545
11	PSOIL	1.281	-0.897	1.314	4.122	36	PHIDE	3.712	1.292	3.769	7.172
12	PSUNO	1.144	-2.583	0.968	5.530	37	PALUM	2.731	-0.747	2.475	5.245
13	PBEEF	1.069	0.010	1.034	2.133	38	PCOPP	0.894	-1.943	0.865	4.835
14	PLAMB	0.646	-0.461	0.677	1.738	39	PIORECR	-1.049	-3.333	-0.943	1.539
15	PPORK	1.043	-2.087	0.944	4.153	40	PLEAD	3.000	0.678	3.092	6.337
16	PPOULT	-0.142	-1.052	-0.147	0.610	41	PNICK	5.009	-0.364	4.642	9.667
17	PFISH	1.053	-0.296	0.975	2.890	42	PTIN	1.020	-0.895	1.146	3.473
18	PSALM	0.054	-1.519	0.174	2.247	43	PURAN	0.773	-1.327	0.838	3.798
19	PSHRI	-0.286	-2.299	-0.238	1.006	44	PZINC	2.849	0.208	2.858	6.467
20	PBANSOP	-2.759	-7.729	-3.074	1.410	45	PCOALAU	2.725	0.880	2.669	5.695
21	PORANG	-0.663	-4.584	-0.907	3.999	46	POILAPSP	1.778	-1.894	1.820	7.053
22	PSUGAISA	-0.827	-4.166	-0.766	2.476	47	POILBRE	1.524	-2.187	1.609	6.931
23	PSUGAUSA	-0.422	-1.624	-0.468	0.665	48	POILDUB	1.746	-2.052	1.793	7.252
24	PCOCO	-1.020	-2.662	-1.041	1.020	49	POILWTI	2.173	-1.505	2.208	7.286
25	PCOFFOTM	1.936	-0.763	1.717	4.959	Mean: 0.853, Median: 0.894					

Note: $p\%$ denotes the p^{th}percentile obtained from 500 nonparametric bootstrap simulations. 50% is the median elasticity estimate. 5% and 95% constitutes the 90% nonparametric confidence band.

Table A.5: Long-Run Dynamic Elasticity with respect to the Real GDP

$$\eta_y^p(\infty) = \frac{\psi_y^p(\infty)}{\psi_y^y(\infty)},\ \psi_y^p(\infty) = \sum_{s=0}^{\infty} \rho_y^p(s),\ \psi_y^y(\infty) = \sum_{s=0}^{\infty} \rho_y^y(s)$$

ID	IMF Code	$\eta_y^p(\infty)$	5%	50%	95%	ID	IMF Code	$\eta_y^p(\infty)$	5%	50%	95%
1	PBARL	0.826	-0.356	0.840	1.988	26	PCOFFORB	-0.115	-1.595	-0.046	1.486
2	PGNUTS	1.200	-0.613	1.226	2.708	27	PTEA	-0.413	-1.474	-0.352	0.736
3	PMAIZMT	0.951	-0.219	0.939	2.175	28	PLOGORE	0.093	-0.670	0.061	0.752
4	PRICENPQ	0.720	-0.761	0.726	1.820	29	PLOGSK	-0.219	-1.498	-0.139	1.047
5	PSMEA	0.200	-1.089	0.255	1.552	30	PSAWMAL	-0.314	-1.645	-0.308	1.012
6	PSOYB	0.463	-0.687	0.434	1.645	31	PSAWORE	-0.063	-0.616	-0.063	0.487
7	PWHEAMT	1.466	0.373	1.542	2.645	32	PCOTTIND	-0.871	-2.602	-0.874	0.725
8	PROIL	0.742	-0.736	0.771	2.161	33	PWOOLC	-0.234	-1.468	-0.129	0.938
9	POLVOIL	-0.757	-1.913	-0.777	0.222	34	PWOOLF	0.290	-1.199	0.283	1.332
10	PPOIL	-0.873	-3.098	-0.837	1.215	35	PRUBB	-0.980	-2.565	-0.903	0.493
11	PSOIL	0.355	-1.077	0.285	1.674	36	PHIDE	-0.329	-1.348	-0.267	0.790
12	PSUNO	0.951	-0.892	0.929	2.623	37	PALUM	-0.426	-2.005	-0.429	0.623
13	PBEEF	0.177	-0.370	0.168	0.706	38	PCOPP	0.354	-1.175	0.382	1.672
14	PLAMB	0.248	-0.555	0.255	1.081	39	PIORECR	-0.134	-1.166	-0.085	0.926
15	PPORK	0.032	-1.497	-0.019	1.440	40	PLEAD	0.706	-1.305	0.770	2.553
16	PPOULT	0.384	0.007	0.381	0.755	41	PNICK	0.451	-2.082	0.429	2.199
17	PFISH	-0.663	-2.014	-0.650	0.572	42	PTIN	1.005	-0.221	1.030	2.401
18	PSALM	-0.352	-1.279	-0.336	0.508	43	PURAN	0.149	-1.686	0.143	1.907
19	PSHRI	0.308	-0.636	0.313	1.147	44	PZINC	0.418	-1.485	0.434	1.851
20	PBANSOP	0.906	-0.404	0.893	1.982	45	PCOALAU	1.921	0.697	1.958	3.276
21	PORANG	0.160	-1.010	0.173	1.551	46	POILAPSP	0.636	-0.911	0.637	2.254
22	PSUGAISA	0.768	-1.172	0.756	2.537	47	POILBRE	0.722	-0.813	0.740	2.384
23	PSUGAUSA	-0.212	-1.166	-0.241	0.595	48	POILDUB	0.559	-0.993	0.538	2.251
24	PCOCO	-0.944	-2.182	-0.930	0.342	49	POILWTI	0.755	-0.673	0.753	2.283
25	PCOFFOTM	0.116	-1.615	0.174	1.810	Mean: 0.227, Median: 0.248					

Note: $p\%$ denotes the p^{th}percentile obtained from 500 nonparametric bootstrap simulations. 50% is the median elasticity estimate. 5% and 95% constitutes the 90% nonparametric confidence band. Long-run responses are obtained by taking the 40^{th}period ahead response function estimate.

Table A.6: Contemporaneous Contribution of the Exchange Rate Shock Relative to the Real GDP Shock

$$\phi(0) = \frac{|\eta_e^p(0)|}{|\eta_e^p(0)|+\left|\eta_y^p(0)\right|}$$

ID	IMF Code	$\phi(0)$	5%	50%	95%	ID	IMF Code	$\phi(0)$	5%	50%	95%
1	PBARL	0.387	0.111	0.356	0.871	26	PCOFFORB	0.598	0.096	0.411	0.871
2	PGNUTS	0.164	0.022	0.165	0.505	27	PTEA	0.565	0.156	0.444	0.922
3	PMAIZMT	0.432	0.020	0.282	0.819	28	PLOGORE	0.322	0.022	0.215	0.735
4	PRICENPQ	0.508	0.016	0.147	0.620	29	PLOGSK	0.348	0.148	0.352	0.828
5	PSMEA	0.785	0.143	0.433	0.861	30	PSAWMAL	0.322	0.135	0.317	0.840
6	PSOYB	0.401	0.116	0.361	0.866	31	PSAWORE	0.945	0.080	0.346	0.819
7	PWHEAMT	0.933	0.183	0.488	0.897	32	PCOTTIND	0.222	0.033	0.218	0.723
8	PROIL	0.202	0.071	0.201	0.468	33	PWOOLC	0.333	0.120	0.318	0.789
9	POLVOIL	0.414	0.244	0.407	0.777	34	PWOOLF	0.132	0.033	0.137	0.361
10	PPOIL	0.449	0.056	0.292	0.828	35	PRUBB	0.345	0.166	0.349	0.830
11	PSOIL	0.261	0.040	0.245	0.710	36	PHIDE	0.058	0.005	0.070	0.186
12	PSUNO	0.255	0.019	0.188	0.767	37	PALUM	0.310	0.164	0.329	0.738
13	PBEEF	0.039	0.007	0.094	0.407	38	PCOPP	0.651	0.257	0.563	0.921
14	PLAMB	0.576	0.318	0.563	0.912	39	PIORECR	0.143	0.010	0.148	0.701
15	PPORK	0.301	0.021	0.238	0.797	40	PLEAD	0.281	0.125	0.271	0.600
16	PPOULT	0.433	0.029	0.246	0.831	41	PNICK	0.173	0.047	0.189	0.650
17	PFISH	0.356	0.132	0.354	0.835	42	PTIN	0.346	0.051	0.297	0.807
18	PSALM	0.958	0.315	0.597	0.945	43	PURAN	0.137	0.019	0.169	0.708
19	PSHRI	0.323	0.017	0.202	0.791	44	PZINC	0.209	0.068	0.205	0.612
20	PBANSOP	0.232	0.032	0.200	0.735	45	PCOALAU	0.155	0.047	0.156	0.366
21	PORANG	0.354	0.016	0.209	0.766	46	POILAPSP	0.368	0.073	0.325	0.858
22	PSUGAISA	0.603	0.178	0.483	0.916	47	POILBRE	0.433	0.094	0.361	0.877
23	PSUGAUSA	0.422	0.100	0.352	0.898	48	POILDUB	0.364	0.063	0.318	0.846
24	PCOCO	0.410	0.170	0.380	0.859	49	POILWTI	0.305	0.054	0.280	0.832
25	PCOFFOTM	0.173	0.027	0.201	0.680	Mean: 0.376, Median: 0.346					

Note: p% denotes the p^{th}percentile obtained from 500 nonparametric bootstrap simulations. 50% is the median elasticity estimate. 5% and 95% constitutes the 90% nonparametric confidence band.

Table A.7: Long-Run Contribution of the Exchange Rate Shock Relative to the Real GDP Shock

$$\phi(\infty) = \frac{|\eta_e^p(\infty)|}{|\eta_e^p(\infty)| + \left|\eta_y^p(\infty)\right|}$$

ID	IMF Code	$\phi(\infty)$	5%	50%	95%	ID	IMF Code	$\phi(\infty)$	5%	50%	95%
1	PBARL	0.657	0.421	0.654	0.942	26	PCOFFORB	0.921	0.246	0.693	0.944
2	PGNUTS	0.639	0.385	0.622	0.918	27	PTEA	0.613	0.062	0.573	0.952
3	PMAIZMT	0.564	0.308	0.570	0.891	28	PLOGORE	0.620	0.054	0.425	0.894
4	PRICENPQ	0.691	0.432	0.665	0.950	29	PLOGSK	0.806	0.195	0.619	0.955
5	PSMEA	0.859	0.339	0.686	0.949	30	PSAWMAL	0.749	0.234	0.622	0.936
6	PSOYB	0.734	0.412	0.678	0.963	31	PSAWORE	0.203	0.038	0.390	0.898
7	PWHEAMT	0.424	0.200	0.425	0.691	32	PCOTTIND	0.476	0.076	0.465	0.914
8	PROIL	0.582	0.139	0.562	0.918	33	PWOOLC	0.811	0.276	0.655	0.948
9	POLVOIL	0.598	0.245	0.589	0.943	34	PWOOLF	0.776	0.232	0.645	0.938
10	PPOIL	0.433	0.036	0.412	0.924	35	PRUBB	0.622	0.319	0.628	0.939
11	PSOIL	0.718	0.197	0.600	0.960	36	PHIDE	0.641	0.094	0.559	0.925
12	PSUNO	0.630	0.256	0.603	0.925	37	PALUM	0.781	0.411	0.729	0.961
13	PBEEF	0.537	0.063	0.495	0.887	38	PCOPP	0.815	0.425	0.705	0.966
14	PLAMB	0.802	0.449	0.720	0.971	39	PIORECR	0.829	0.136	0.605	0.950
15	PPORK	0.581	0.052	0.395	0.895	40	PLEAD	0.621	0.146	0.529	0.909
16	PPOULT	0.304	0.049	0.313	0.698	41	PNICK	0.777	0.236	0.638	0.944
17	PFISH	0.417	0.038	0.383	0.912	42	PTIN	0.404	0.057	0.418	0.886
18	PSALM	0.693	0.214	0.661	0.962	43	PURAN	0.877	0.142	0.564	0.936
19	PSHRI	0.437	0.064	0.418	0.901	44	PZINC	0.580	0.078	0.487	0.883
20	PBANSOP	0.293	0.04	0.333	0.814	45	PCOALAU	0.506	0.339	0.504	0.714
21	PORANG	0.737	0.072	0.498	0.914	46	POILAPSP	0.588	0.123	0.521	0.920
22	PSUGAISA	0.690	0.300	0.617	0.953	47	POILBRE	0.579	0.147	0.514	0.929
23	PSUGAUSA	0.674	0.089	0.531	0.930	48	POILDUB	0.607	0.093	0.517	0.885
24	PCOCO	0.018	0.040	0.280	0.779	49	POILWTI	0.533	0.113	0.485	0.928
25	PCOFFOTM	0.908	0.180	0.625	0.946	Mean: 0.619, Median: 0.622					

Note: p% denotes the p^{th}percentile obtained from 500 nonparametric bootstrap simulations. 50% is the median elasticity estimate. 5% and 95% constitutes the 90% nonparametric confidence band. Long-run responses are obtained by taking the 40^{th}period ahead response function estimate.

Table A.8: Contemporaneous Dynamic Elasticity with respect to the Exchange Rate: Real Commodity Prices

$$\eta_e^p(0) = \frac{\psi_e^p(0)}{\psi_e^e(0)},\ \psi_e^p(0) = \rho_e^p(0),\ \psi_e^e(0) = \rho_e^e(0)$$

ID	IMF Code	$\eta_e^p(0)$	5%	50%	95%	ID	IMF Code	$\eta_e^p(0)$	5%	50%	95%
1	PBARL	-0.819	-1.389	-0.852	-0.304	26	PCOFFORB	-0.633	-1.214	-0.641	0.023
2	PGNUTS	-0.704	-1.412	-0.733	-0.113	27	PTEA	-0.824	-1.339	-0.816	-0.328
3	PMAIZMT	-0.288	-0.774	-0.321	0.193	28	PLOGORE	-0.047	-0.378	-0.049	0.229
4	PRICENPQ	-0.224	-0.659	-0.233	0.192	29	PLOGSK	-0.779	-1.219	-0.786	-0.350
5	PSMEA	-0.685	-1.179	-0.696	-0.247	30	PSAWMAL	-0.670	-1.026	-0.656	-0.335
6	PSOYB	-0.667	-1.138	-0.676	-0.231	31	PSAWORE	0.322	0.072	0.315	0.592
7	PWHEAMT	-0.770	-1.304	-0.752	-0.328	32	PCOTTIND	-0.359	-0.810	-0.389	0.140
8	PROIL	-0.933	-1.495	-0.945	-0.426	33	PWOOLC	-0.644	-1.059	-0.666	-0.216
9	POLVOIL	-1.126	-1.448	-1.133	-0.832	34	PWOOLF	-0.592	-1.010	-0.616	-0.173
10	PPOIL	-0.731	-1.340	-0.751	-0.164	35	PRUBB	-1.395	-2.081	-1.366	-0.799
11	PSOIL	-0.531	-1.072	-0.555	-0.073	36	PHIDE	-0.405	-1.077	-0.378	0.242
12	PSUNO	-0.468	-1.118	-0.481	0.093	37	PALUM	-1.223	-1.678	-1.226	-0.719
13	PBEEF	-0.042	-0.282	-0.048	0.178	38	PCOPP	-1.728	-2.383	-1.705	-1.065
14	PLAMB	-0.819	-1.081	-0.821	-0.554	39	PIORECR	0.228	-0.241	0.214	0.612
15	PPORK	0.354	-0.318	0.358	1.044	40	PLEAD	-1.088	-1.796	-1.106	-0.416
16	PPOULT	0.120	-0.060	0.114	0.274	41	PNICK	-1.055	-1.818	-1.048	-0.277
17	PFISH	-0.509	-0.832	-0.515	-0.151	42	PTIN	-0.556	-1.171	-0.579	0.024
18	PSALM	-1.195	-1.549	-1.221	-0.857	43	PURAN	0.003	-0.445	-0.015	0.491
19	PSHRI	-0.047	-0.359	-0.060	0.229	44	PZINC	-0.755	-1.419	-0.793	-0.180
20	PBANSOP	0.699	-0.319	0.645	1.497	45	PCOALAU	-0.693	-1.194	-0.695	-0.236
21	PORANG	-0.447	-1.355	-0.462	0.483	46	POILAPSP	-1.451	-2.370	-1.450	-0.593
22	PSUGAISA	-1.080	-1.781	-1.105	-0.403	47	POILBRE	-1.587	-2.493	-1.565	-0.724
23	PSUGAUSA	-0.274	-0.498	-0.278	-0.048	48	POILDUB	-1.409	-2.406	-1.383	-0.547
24	PCOCO	-0.666	-1.050	-0.686	-0.302	49	POILWTI	-1.341	-2.289	-1.336	-0.481
25	PCOFFOTM	-0.428	-1.136	-0.427	0.255	Mean: -0.632, Median:-0.667					

Note: We divided nominal commodity prices by the US consumer price index to get real commodity prices, because nominal commodity prices are denominated in US dollars. $p\%$ denotes the p^{th}percentile obtained from 500 nonparametric bootstrap simulations. 50% is the median elasticity estimate. 5% and 95% constitutes the 90% nonparametric confidence band.

Table A.9: Long-Run Dynamic Elasticity with respect to the Exchange Rate: Real Commodity Prices

$$\eta_e^p(\infty) = \frac{\psi_e^p(\infty)}{\psi_e^e(\infty)},\ \psi_e^p(\infty) = \sum_{s=0}^{\infty} \rho_e^p(s),\ \psi_e^e(\infty) = \sum_{s=0}^{\infty} \rho_e^e(s)$$

ID	IMF Code	$\eta_e^p(\infty)$	5%	50%	95%	ID	IMF Code	$\eta_e^p(\infty)$	5%	50%	95%
1	PBARL	-1.527	-2.481	-1.523	-0.719	26	PCOFFORB	-1.190	-2.145	-1.225	-0.282
2	PGNUTS	-2.082	-3.301	-2.103	-1.012	27	PTEA	-0.635	-1.595	-0.654	0.175
3	PMAIZMT	-1.168	-2.104	-1.187	-0.365	28	PLOGORE	-0.158	-0.658	-0.131	0.330
4	PRICENPQ	-1.600	-2.348	-1.607	-0.845	29	PLOGSK	-0.796	-1.593	-0.812	-0.012
5	PSMEA	-1.006	-1.863	-1.010	-0.195	30	PSAWMAL	-0.871	-1.532	-0.850	-0.262
6	PSOYB	-1.118	-1.929	-1.128	-0.376	31	PSAWORE	0.021	-0.309	0.019	0.383
7	PWHEAMT	-1.031	-1.928	-1.027	-0.309	32	PCOTTIND	-0.629	-1.508	-0.692	0.284
8	PROIL	-0.935	-1.939	-0.955	0.054	33	PWOOLC	-0.796	-1.481	-0.794	-0.084
9	POLVOIL	-1.047	-1.730	-1.050	-0.426	34	PWOOLF	-0.950	-1.713	-0.961	-0.219
10	PPOIL	-0.497	-1.699	-0.518	0.660	35	PRUBB	-1.508	-2.176	-1.500	-0.827
11	PSOIL	-0.786	-1.722	-0.810	0.074	36	PHIDE	-0.614	-1.290	-0.577	0.060
12	PSUNO	-1.660	-2.985	-1.728	-0.462	37	PALUM	-1.361	-1.997	-1.363	-0.748
13	PBEEF	-0.125	-0.496	-0.143	0.258	38	PCOPP	-1.445	-2.370	-1.406	-0.521
14	PLAMB	-0.898	-1.453	-0.880	-0.385	39	PIORECR	-0.563	-1.286	-0.559	0.127
15	PPORK	-0.113	-0.986	-0.092	0.841	40	PLEAD	-0.794	-1.950	-0.768	0.525
16	PPOULT	-0.090	-0.365	-0.088	0.161	41	PNICK	-1.317	-2.590	-1.331	-0.162
17	PFISH	-0.274	-1.033	-0.268	0.572	42	PTIN	-0.462	-1.660	-0.493	0.614
18	PSALM	-0.617	-1.148	-0.606	-0.027	43	PURAN	-0.739	-1.839	-0.751	0.463
19	PSHRI	-0.162	-0.762	-0.203	0.432	44	PZINC	-0.424	-1.491	-0.437	0.706
20	PBANSOP	-0.378	-1.167	-0.360	0.382	45	PCOALAU	-2.094	-3.174	-2.074	-1.146
21	PORANG	-0.442	-1.273	-0.449	0.491	46	POILAPSP	-1.076	-2.179	-1.054	-0.011
22	PSUGAISA	-1.308	-2.487	-1.268	-0.050	47	POILBRE	-1.173	-2.329	-1.135	-0.106
23	PSUGAUSA	-0.253	-0.801	-0.250	0.288	48	POILDUB	-1.033	-2.148	-0.998	0.026
24	PCOCO	0.182	-0.674	0.163	1.064	49	POILWTI	-1.006	-2.183	-0.974	-0.038
25	PCOFFOTM	-1.084	-2.207	-1.139	0.047	Mean: -0.850, Median:-0.871					

Note: We divided nominal commodity prices by the US consumer price index to get real commodity prices, because nominal commodity prices are denominated in US dollars. $p\%$ denotes the p^{th}percentile obtained from 500 nonparametric bootstrap simulations. 50% is the median elasticity estimate. 5% and 95% constitutes the 90% nonparametric confidence band. Long-run responses are obtained by taking the 40^{th}period ahead response function estimate.

Table A.10: Contemporaneous Dynamic Elasticity with respect to the Real GDP: Real Commodity Prices

$$\eta_y^p(0) = \frac{\psi_y^p(0)}{\psi_y^y(0)},\ \psi_y^p(0) = \rho_y^p(0),\ \psi_y^y(0) = \rho_y^y(0)$$

ID	IMF Code	$\eta_y^p(0)$	5%	50%	95%	ID	IMF Code	$\eta_y^p(0)$	5%	50%	95%
1	PBARL	1.057	-1.066	1.075	3.714	26	PCOFFORB	0.461	-1.719	0.282	2.840
2	PGNUTS	3.266	0.878	3.330	6.065	27	PTEA	-0.799	-3.342	-0.860	1.416
3	PMAIZMT	0.248	-1.486	0.218	2.099	28	PLOGORE	-0.323	-1.985	-0.262	0.943
4	PRICENPQ	-0.178	-3.911	-0.266	2.264	29	PLOGSK	-1.696	-4.311	-1.513	1.151
5	PSMEA	0.251	-1.617	0.264	3.111	30	PSAWMAL	-1.570	-4.173	-1.278	1.33
6	PSOYB	0.996	-0.779	1.022	3.391	31	PSAWORE	-0.042	-1.632	-0.036	1.156
7	PWHEAMT	-0.115	-1.815	-0.171	1.964	32	PCOTTIND	-1.084	-2.993	-1.047	1.567
8	PROIL	3.614	1.328	3.710	6.227	33	PWOOLC	1.310	-0.247	1.395	3.350
9	POLVOIL	-1.752	-3.389	-1.780	-0.510	34	PWOOLF	3.357	0.852	3.304	5.229
10	PPOIL	0.827	-1.953	0.804	5.135	35	PRUBB	2.127	-0.357	2.109	5.998
11	PSOIL	1.308	-0.721	1.365	3.956	36	PHIDE	3.604	1.331	3.703	6.935
12	PSUNO	1.098	-2.445	0.948	5.218	37	PALUM	2.454	-0.859	2.229	4.995
13	PBEEF	1.079	-0.010	1.033	2.197	38	PCOPP	0.750	-1.940	0.695	4.378
14	PLAMB	0.624	-0.419	0.654	1.651	39	PIORECR	-1.095	-3.402	-0.988	1.528
15	PPORK	0.985	-2.007	0.890	3.990	40	PLEAD	2.985	0.844	3.054	6.183
16	PPOULT	-0.084	-0.942	-0.111	0.662	41	PNICK	4.994	-0.458	4.655	9.698
17	PFISH	1.034	-0.377	0.947	2.865	42	PTIN	1.059	-0.761	1.172	3.403
18	PSALM	0.038	-1.559	0.124	2.251	43	PURAN	0.855	-1.272	0.942	3.899
19	PSHRI	-0.338	-2.415	-0.305	0.963	44	PZINC	2.804	0.255	2.804	6.346
20	PBANSOP	-2.770	-8.002	-3.091	1.439	45	PCOALAU	2.600	0.881	2.579	5.265
21	PORANG	-0.762	-4.687	-0.914	3.771	46	POILAPSP	1.499	-1.900	1.531	6.288
22	PSUGAISA	-0.791	-4.233	-0.740	2.438	47	POILBRE	1.230	-2.242	1.312	6.188
23	PSUGAUSA	-0.413	-1.584	-0.449	0.621	48	POILDUB	1.467	-2.199	1.551	6.472
24	PCOCO	-1.037	-2.611	-1.077	0.912	49	POILWTI	1.948	-1.316	1.954	6.542
25	PCOFFOTM	1.857	-0.887	1.675	4.943		Mean: 0.795, Median: 0.855				

Note: We divided nominal commodity prices by the US consumer price index to get real commodity prices, because nominal commodity prices are denominated in US dollars. p% denotes the p^{th}percentile obtained from 500 nonparametric bootstrap simulations. 50% is the median elasticity estimate. 5% and 95% constitutes the 90% nonparametric confidence band.

Table A.11: Long-Run Dynamic Elasticity with respect to the Real GDP: Real Commodity Prices

$$\eta_y^p(\infty) = \frac{\psi_y^p(\infty)}{\psi_y^y(\infty)},\ \psi_y^p(\infty) = \sum_{s=0}^{\infty} \rho_y^p(s),\ \psi_y^y(\infty) = \sum_{s=0}^{\infty} \rho_y^y(s)$$

ID	IMF Code	$\eta_y^p(\infty)$	5%	50%	95%	ID	IMF Code	$\eta_y^p(\infty)$	5%	50%	95%
1	PBARL	0.817	-0.303	0.837	1.96	26	PCOFFORB	-0.128	-1.592	-0.041	1.459
2	PGNUTS	1.177	-0.653	1.145	2.694	27	PTEA	-0.468	-1.542	-0.388	0.612
3	PMAIZMT	0.939	-0.243	0.918	2.144	28	PLOGORE	0.035	-0.698	0.003	0.695
4	PRICENPQ	0.696	-0.828	0.714	1.772	29	PLOGSK	-0.220	-1.530	-0.158	1.053
5	PSMEA	0.222	-1.141	0.249	1.571	30	PSAWMAL	-0.341	-1.685	-0.339	0.971
6	PSOYB	0.479	-0.682	0.439	1.639	31	PSAWORE	-0.096	-0.666	-0.089	0.456
7	PWHEAMT	1.449	0.421	1.502	2.545	32	PCOTTIND	-0.831	-2.537	-0.843	0.652
8	PROIL	0.716	-0.729	0.724	2.106	33	PWOOLC	-0.249	-1.466	-0.153	0.972
9	POLVOIL	-0.811	-1.954	-0.819	0.144	34	PWOOLF	0.265	-1.216	0.257	1.302
10	PPOIL	-0.859	-3.030	-0.857	1.137	35	PRUBB	-1.043	-2.682	-1.003	0.440
11	PSOIL	0.366	-1.049	0.280	1.619	36	PHIDE	-0.382	-1.414	-0.322	0.688
12	PSUNO	0.942	-0.817	0.914	2.697	37	PALUM	-0.552	-2.127	-0.607	0.563
13	PBEEF	0.160	-0.386	0.154	0.703	38	PCOPP	0.272	-1.227	0.280	1.561
14	PLAMB	0.229	-0.592	0.226	1.082	39	PIORECR	-0.170	-1.184	-0.116	0.872
15	PPORK	-0.006	-1.525	-0.037	1.392	40	PLEAD	0.721	-1.201	0.782	2.527
16	PPOULT	0.379	-0.021	0.375	0.781	41	PNICK	0.425	-2.275	0.352	2.219
17	PFISH	-0.703	-2.096	-0.709	0.566	42	PTIN	1.021	-0.233	1.039	2.383
18	PSALM	-0.353	-1.305	-0.338	0.450	43	PURAN	0.187	-1.643	0.201	1.904
19	PSHRI	0.268	-0.671	0.258	1.115	44	PZINC	0.390	-1.444	0.409	1.807
20	PBANSOP	0.884	-0.364	0.851	1.955	45	PCOALAU	1.881	0.701	1.926	3.174
21	PORANG	0.126	-1.037	0.142	1.469	46	POILAPSP	0.576	-0.949	0.549	2.160
22	PSUGAISA	0.803	-1.238	0.795	2.529	47	POILBRE	0.659	-0.902	0.680	2.298
23	PSUGAUSA	-0.197	-1.185	-0.232	0.639	48	POILDUB	0.501	-0.972	0.483	2.123
24	PCOCO	-0.964	-2.176	-0.952	0.332	49	POILWTI	0.702	-0.708	0.680	2.184
25	PCOFFOTM	0.078	-1.621	0.118	1.827	Mean: 0.204, Median: 0.229					

Note: We divided nominal commodity prices by the US consumer price index to get real commodity prices, because nominal commodity prices are denominated in US dollars. $p\%$ denotes the p^{th}percentile obtained from 500 nonparametric bootstrap simulations. 50% is the median elasticity estimate. 5% and 95% constitutes the 90% nonparametric confidence band. Long-run responses are obtained by taking the 40^{th}period ahead response function estimate.

Table A.12: Contemporaneous Contribution of the Exchange Rate Shock Relative to the Real GDP Shock: Real Commodity Prices

$$\phi(0) = \frac{|\eta_e^p(0)|}{|\eta_e^p(0)| + |\eta_y^p(0)|}$$

ID	IMF Code	$\phi(0)$	5%	50%	95%	ID	IMF Code	$\phi(0)$	5%	50%	95%
1	PBARL	0.437	0.142	0.398	0.853	26	PCOFFORB	0.579	0.076	0.383	0.896
2	PGNUTS	0.177	0.028	0.175	0.529	27	PTEA	0.508	0.153	0.420	0.894
3	PMAIZMT	0.537	0.037	0.310	0.869	28	PLOGORE	0.128	0.013	0.178	0.706
4	PRICENPQ	0.557	0.026	0.177	0.700	29	PLOGSK	0.315	0.116	0.318	0.826
5	PSMEA	0.732	0.148	0.427	0.898	30	PSAWMAL	0.299	0.122	0.298	0.820
6	PSOYB	0.401	0.129	0.379	0.857	31	PSAWORE	0.884	0.093	0.357	0.850
7	PWHEAMT	0.870	0.189	0.519	0.925	32	PCOTTIND	0.249	0.037	0.238	0.789
8	PROIL	0.205	0.080	0.203	0.455	33	PWOOLC	0.329	0.110	0.313	0.823
9	POLVOIL	0.391	0.231	0.386	0.686	34	PWOOLF	0.150	0.043	0.158	0.379
10	PPOIL	0.469	0.086	0.333	0.824	35	PRUBB	0.396	0.199	0.394	0.820
11	PSOIL	0.289	0.061	0.273	0.801	36	PHIDE	0.101	0.010	0.099	0.253
12	PSUNO	0.299	0.026	0.228	0.819	37	PALUM	0.333	0.161	0.348	0.809
13	PBEEF	0.037	0.009	0.087	0.409	38	PCOPP	0.697	0.282	0.567	0.943
14	PLAMB	0.568	0.303	0.547	0.908	39	PIORECR	0.172	0.014	0.168	0.729
15	PPORK	0.265	0.027	0.215	0.787	40	PLEAD	0.267	0.111	0.256	0.555
16	PPOULT	0.590	0.026	0.273	0.806	41	PNICK	0.174	0.044	0.189	0.653
17	PFISH	0.330	0.097	0.329	0.836	42	PTIN	0.344	0.051	0.304	0.818
18	PSALM	0.969	0.302	0.605	0.945	43	PURAN	0.003	0.015	0.145	0.605
19	PSHRI	0.121	0.015	0.147	0.720	44	PZINC	0.212	0.065	0.206	0.596
20	PBANSOP	0.201	0.022	0.181	0.716	45	PCOALAU	0.210	0.077	0.210	0.431
21	PORANG	0.370	0.023	0.211	0.784	46	POILAPSP	0.492	0.143	0.427	0.903
22	PSUGAISA	0.577	0.145	0.453	0.914	47	POILBRE	0.564	0.173	0.468	0.910
23	PSUGAUSA	0.399	0.073	0.330	0.822	48	POILDUB	0.490	0.128	0.414	0.905
24	PCOCO	0.391	0.162	0.364	0.866	49	POILWTI	0.408	0.130	0.378	0.874
25	PCOFFOTM	0.187	0.026	0.203	0.737	Mean: 0.381, Median: 0.344					

Note: We divided nominal commodity prices by the US consumer price index to get real commodity prices, because nominal commodity prices are denominated in US dollars. $p\%$ denotes the p^{th}percentile obtained from 500 nonparametric bootstrap simulations. 50% is the median elasticity estimate. 5% and 95% constitutes the 90% nonparametric confidence band.

Table A.13: Long-Run Contribution of the Exchange Rate Shock Relative to the Real GDP Shock: Real Commodity Prices

$$\phi(\infty) = \frac{|\eta_e^p(\infty)|}{|\eta_e^p(\infty)|+\left|\eta_y^p(\infty)\right|}$$

ID	IMF Code	$\phi(\infty)$	5%	50%	95%	ID	IMF Code	$\phi(\infty)$	5%	50%	95%
1	PBARL	0.651	0.416	0.656	0.937	26	PCOFFORB	0.903	0.189	0.667	0.949
2	PGNUTS	0.639	0.391	0.622	0.932	27	PTEA	0.576	0.064	0.556	0.947
3	PMAIZMT	0.554	0.288	0.555	0.908	28	PLOGORE	0.816	0.047	0.431	0.912
4	PRICENPQ	0.697	0.440	0.666	0.958	29	PLOGSK	0.783	0.130	0.583	0.937
5	PSMEA	0.819	0.239	0.645	0.957	30	PSAWMAL	0.719	0.202	0.593	0.943
6	PSOYB	0.700	0.348	0.649	0.954	31	PSAWORE	0.181	0.042	0.371	0.916
7	PWHEAMT	0.416	0.176	0.413	0.674	32	PCOTTIND	0.431	0.058	0.425	0.888
8	PROIL	0.566	0.129	0.532	0.898	33	PWOOLC	0.762	0.161	0.611	0.955
9	POLVOIL	0.564	0.213	0.551	0.938	34	PWOOLF	0.782	0.244	0.626	0.931
10	PPOIL	0.366	0.043	0.379	0.898	35	PRUBB	0.591	0.297	0.595	0.938
11	PSOIL	0.683	0.147	0.573	0.948	36	PHIDE	0.617	0.102	0.561	0.931
12	PSUNO	0.638	0.275	0.604	0.938	37	PALUM	0.711	0.349	0.671	0.945
13	PBEEF	0.439	0.058	0.433	0.868	38	PCOPP	0.842	0.356	0.689	0.953
14	PLAMB	0.797	0.415	0.705	0.955	39	PIORECR	0.768	0.115	0.557	0.930
15	PPORK	0.953	0.037	0.392	0.887	40	PLEAD	0.524	0.074	0.455	0.897
16	PPOULT	0.191	0.040	0.255	0.671	41	PNICK	0.756	0.183	0.590	0.917
17	PFISH	0.281	0.039	0.325	0.886	42	PTIN	0.312	0.050	0.374	0.801
18	PSALM	0.636	0.123	0.597	0.937	43	PURAN	0.798	0.102	0.488	0.898
19	PSHRI	0.376	0.072	0.421	0.870	44	PZINC	0.521	0.060	0.460	0.901
20	PBANSOP	0.299	0.036	0.332	0.810	45	PCOALAU	0.527	0.361	0.522	0.738
21	PORANG	0.778	0.059	0.511	0.925	46	POILAPSP	0.651	0.163	0.577	0.920
22	PSUGAISA	0.619	0.189	0.540	0.933	47	POILBRE	0.640	0.193	0.574	0.935
23	PSUGAUSA	0.562	0.074	0.447	0.906	48	POILDUB	0.673	0.137	0.570	0.902
24	PCOCO	0.159	0.030	0.289	0.755	49	POILWTI	0.589	0.158	0.544	0.923
25	PCOFFOTM	0.933	0.165	0.615	0.949	Mean: 0.608, Median: 0.638					

Note: We divided nominal commodity prices by the US consumer price index to get real commodity prices, because nominal commodity prices are denominated in US dollars. $p\%$ denotes the p^{th}percentile obtained from 500 nonparametric bootstrap simulations. 50% is the median elasticity estimate. 5% and 95% constitutes the 90% nonparametric confidence band. Long-run responses are obtained by taking the 40^{th}period ahead response function estimate.