Three Essays in Applied Economics

by

Matthew K. Muntifering

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 8, 2020

Keywords: Price elasticity, Retail energy demands, Asset pricing models, Investor sentiment, Thoroughbred broodmare auctions

Approved by

Henry Kinnucan, Chair, Alumni Professor, Agricultural Economics and Rural Sociology.
Henry Thompson, Professor Emeritus, Economics.
Ruiqing Miao, Assistant Professor, Agricultural Economics and Rural Sociology.
John Ng'ombe, Visiting Assistant Professor, Agricultural Economics and Rural Sociology.

Abstract

The dissertation begins with a brief discussion of the benefits of using an essay style approach compared to a traditional opus in chapter one. Chapter two examines the consumers facing the implications of energy regulation policies, United States residents, who have increasing control over their energy consumption in response to price changes. This chapter paper estimates the price elasticity of demand for biomass, distillate fuel oil, hydrocarbon gas liquid, and natural gas using the Exact Affine Stone Index (EASI) demand system, while contributing methodologically the Differential Exact Affine Stone Index (DEASI) demand system.

Chapter three extends the asset pricing literature by offering a proprietary index of negative investor sentiment linked to carbon monoxide (CO), nitrogen dioxide (NO2), ozone particle (O3), 2.5 mm particulate matter (PM2.5), and sulfur dioxide (SO2) levels. Food products and wholesale portfolio returns increase with negative investor sentiment, consistent with psychological traits linked to binge eating and shopping sprees when individuals experience stress. Personal services portfolio returns decrease when negative investor sentiment increases, consistent with the behavior isolationism

Chapter four addresses the issue with OLS estimation in hedonic pricing model literature of not accounting for sample selection bias. In broodmare auctions, the purchased decision and whether a price is realized or zero is endogenous. This chapter contributes to the hedonic broodmare price analysis literature by implementing the Heckman (1976) procedure to control for selection in estimating a hedonic pricing model using data comes from the 2020 January Keenland Sale. A list of published papers does not accommodate this selection process and has biased coefficients.

The sire's stud fee, domestic status, and the day of the session are significant for broodmare prices. This may be implemented within a profit maximizing purchasing and breeding strategy.

Acknowledgements

First, I would like to thank my family and my faith.

I am extremely thankful for Dr. Henry Kinnucan for leading the committee, and showing me the ropes of being a researcher, Dr. Henry Thompson for his insights into energy markets, Dr. Ruiqing Miao for teaching me the joy in reading research papers, Dr. John Ng'ombe for teaching consistency during the coronavirus pandemic, Dr. Emir Malikov for his ability to make econometrics understandable, Dr. Aditi Sengupta for agreeing to serve as the University Reader, and Dr. Joel Cuffey for his recurring motivation and detailed writing tips. I am also thankful for Dr. Patricia Duffy and Dr. James Novak for allowing me to serve as a graduate assistant and course instructor. I thank Dr. Deacue Fields, Dr. Patricia Duffy, and Dr. Joseph Duke for their leadership as department heads during my journey. Thank you Dr. Kinnucan, Dr. Malikov, Dr. Alan Seals, Dr. Duha Altindag, Dr. Randy Beard, Chip Emery, Darron Bradham, Zach Alsobrook, and Roger Bell for believing in me from the very start.

Finally, I am thankful for my classmates, and the faculty and staff in our department who made this process enjoyable by their persistent positive attitudes.

Table of Contents

Abstract…………………………………………………………………………………...........1

Acknowledgments……………………………………...…………………………………….3

List of Tables…………………………………………...…………………………………6

Lift of Figures..…………………………..…………………..………………………………7

Chapter 1: A Brief Discussion on Essay-Style Dissertations

i. Background…………………….……………………………..…………………………..8
ii. References………………………….…………………….....……………………………9

Chapter 2: An EASI Model of U.S. Residential Energy Demand

i. Abstract…………………………………...……………...………………………………11
ii. Keywords…………………………………...…………...…………………………….11
iii. Introduction……………………………………...………...……………………………..11
iv. Methodology…………………………………...……...……………………………13
v. Data and Descriptive Statistics…………………………...………………….....………18
vi. Results and Discussion……………………………………...……...………………19
vii. Conclusion………………………………………...……………...………………...22
viii. References……………………………………...…………………………………...24
ix. Appendix 1: Tables……………………………………...……………….....…….30
x. Appendix 2: Figures……………………………………...…………………...…34

Chapter 3: Air Pollution, Investor Sentiment, and Excessive Returns

i. Abstract……………………………………...…………………………………...35
ii. Keywords……………………………………...…………………………………35
iii. Introduction……………………………………...………………………………….35
iv. Review of Literature……………………………………...……………………...36

v. Theoretical Framework……………………………………………...…………………...39
vi. Data……………………………………..……………………………………….41
vii. Empirical Framework……………………………………...……………………42
viii. Results and Discussion……………………………………...………………..…43
ix. Conclusion……………………………………..……………………….........43
x. References……………………………………..……………………….........45
xi. Appendix 1: Tables……………………………………...………………………50
xii. Appendix 2: Figures……………………………………...…………………...…54

Chapter 4: Sample Selection Bias in Hedonic Pricing Models of Thoroughbred Broodmares

i. Abstract……………………………………..……………………….........57
ii. Keywords……………………………………...…………………………………57
iii. Introduction……………………………………...………………………………57
iv. Methodology……………………………………...……………………………...59
v. Data and Descriptive Statistics……………………………………...…………...60
vi. Results and Discussion……………………………………...…………………...61
vii. Conclusion……………………………………...………………………………..62
viii. References……………………………………...………………………………...63
ix. Appendix 1: Tables……………………………………...………………………66
x. Appendix 2: Figures……………………………………...………………………67

List of Tables

Chapter 2: An EASI Model of U.S. Residential Energy Demand

Table 1: Descriptive Statistics.…………………………………………..…………..………30

Table 2: Results from Restricted EASI from Equation 2, r=3………………………………...30

Table 3: Results from DEASI ………………………………………..…………………….31

Table 4: Marshallian Own-Price Elasticities by Specification………………………………..31

Table 5: Expenditure Elasticities by Specification……………………………………...…….32

Table 6: Hicksian Own-Price Elasticities by Specification…………………………………...32

Table 7: Allen Substitution Elasticities by Specification……………………………………..…33

Chapter 3: Air Pollution, Investor Sentiment, and Excessive Returns

Table 1: Descriptive Statistics.…………………………………...……...………………50

Table 2: Proportion of Variance Explained by Components………………...…………………..51

Table 3: Loading Table……………………………………...…………...………………51

Table 4: Regression Results……………………………………...…………...…………51

Chapter 4: Sample Selection Bias in Hedonic Pricing Models of Thoroughbred Broodmares

Table 1: Descriptive Statistics………………………………………..……………..……..66

Table 2: Pairwise comparisons of mean price across sessions with Bonferroni adjustment…....66

Table 3: Bayesian Heckman Posterior Summary Statistics……………………………………..66

List of Figures

Chapter 2: An EASI Model of U.S. Residential Energy Demand

Figure 1: Average Budget Shares Over Time……………………………….………………...34

Figure 2: Average Budget Shares…………………………………………..........…………...34

Chapter 3: Air Pollution, Investor Sentiment, and Excessive Returns

Figure 1: Carbon Monoxide……………………………………...………………………54

Figure 2: Nitrogen Dioxide……………………………………...……………………….54

Figure 3: Ozone……………………………………..........…………………………….…...54

Figure 4: PM2.5……………………………………...………...………………………...55

Figure 5: Sulfur Dioxide……………………………………........……………………… 55

Figure 6: Scree Plot……………………………………...……………………………….55

Figure 7: Loading Plot……………………………………...………………...………….56

Chapter 4: Sample Selection Bias in Hedonic Pricing Models of Thoroughbred Broodmares

Figure 1: Price of non-prospects vs. prospects……………………………………...……...67

Figure 2: Price difference across Sire's domestic status…………………………………....…67

Figure 3: Price over session……………………………………...………………………68

Figure 4: Diagnostic Plots of Sire's stud fee…………………………………………….....68

Figure 5: Diagnostic Plots of Sire's domestic status……………………………………….69

Chapter 1: A Brief Discussion on Essay-Style Dissertations

Background

Stock and Siegfried (2013) provide a thorough analysis of the historical style of economics dissertations. They note that until around the 1970's that dissertations were mostly on one general subject rather than a series of essays. The essay format lowers a student's hurdle in publication, as they have already produced journal style manuscripts, potentially at a lower cost than deducting separate articles from one composition. Stock and Siegfried (2013) further point out that in 2010, around 69 percent of all dissertations in economics were of the essay format, ranging from two to four, with three being the most popular essay number. They note four main advantages from applying the essay strategy. They are question identification, opportunity cost reduction, career acceleration, practical preparation, and rigorous feedback. Students may be more adapt at finding specific questions, particularly in today's data-rich research environment, than at specifying a general hypothesis over a book-length study. Stock and Siegfried (2006) found that students who write the essay-style dissertations finish the program around 6 months faster on average than those who write in the traditional format. In addition, the time saving translates to an earlier launching of the scientist's career. This style of writing is consistent with the manta *publish or perish,* which the prospect will face in their academic career. Finally, if a manuscript is submitted to a journal, it will undergo the peer-review process in addition to feedback from the dissertation committee, reinforcing the career practicality benefit, and providing a greater scientific rigor to be applied to the studies. Duke and Beck (1999) find that one of the essays from a set within the new style of dissertation research is more likely to get published than a manuscript derived from a traditionally formatted dissertation.

There are arguments to be made against the new approach. One is that PhD programs are meant to train scholars in depth of a field, rather than shallower knowledge across a wider breadth of areas. This is valid but is complemented by even further benefit. A scholar with a wider range of publication areas may attend more meetings, teach more classes, and work with a wider set of researchers than one with a narrower focus. For the reasons listed above, this dissertation will follow the essay style.

References

Duke, Nell K., and Sarah W. Beck. "Research news and comment: Education should consider alternative formats for the dissertation." *Educational Researcher* 28, no. 3 (1999): 31-36.

Stock, Wendy A., and John J. Siegfried. "Time-to-degree for the economics Ph. D. class of 2001-2002." *American Economic Review* 96, no. 2 (2006): 467-474.

Stock, Wendy A., and John J. Siegfried. "One essay on dissertation formats in economics." *American Economic Review* 103, no. 3 (2013): 648-53.

Chapter 2: An EASI Model of U.S. Residential Energy Demand

Abstract

This paper estimates the price elasticity of demand for fuel oil, hydrocarbon gas liquids (HGL), natural gas, and wood using the Exact Affine Stone Index (EASI) demand system, while contributing methodologically the Differential Exact Affine Stone Index (DEASI) demand system.

Keywords

Price elasticity, Retail energy demands, United States, Panel data

Introduction

This paper identifies the demand elasticities for four primary energy sources: distillate fuel oil, hydrocarbon gas liquids, natural gas, and wood in the United States residential sector using the Exact Affine Stone Index (EASI) implicit Marshallian demand system of Lewbel and Pendakur (2009).

According to the EIA, the residential sector is responsible for approximately 7% of the primary energy consumption in the United States. Appliances, amenities, geographic characteristics, number of residents, and fuel-type used affect the level of energy consumption within each household. Space heating and air conditioning account for roughly half of the total household energy consumption; while water heating, lighting, and refrigeration together are roughly a quarter. The remainder is household cooking and cleaning appliances, as well as personal electronic devices. Energy-use efficiency can be increased by improving insulation technology, as well as keeping household appliances current.

Petroleum consumption is via fuel oil, and hydrocarbon gas liquids (HGL) like kerosene and propane. New York, Pennsylvania, Maryland, Connecticut, and Maine were the largest consumers of heating oil in 2018. Roughly 85% of residential heating oil sales came from Northeastern states in the same year. The sensitivity of heating oil demand to weather makes controlling for temperature important. Propane is mostly used for space and water heating, cooking, and drying. Natural gas, also commonly used for space and water heating, cooking, and drying, is used in over half of the homes in the United States. It accounts for roughly 17% of the country's natural gas consumption. Like heating oil, natural gas demand is nationwide but largely consolidated mostly to five states. Texas, California, Louisiana, Florida, and Pennsylvania were responsible for roughly 37% of total natural gas consumption in the United States. Biomass is consumed via biomass waste, biofuels, wood, and wood waste. It is responsible for roughly 45% of the renewable energy consumption in the United States.

The Renewable Fuel Standard (RFS), enacted into legislation in its current form in 2007, was designed to mitigate the economic cost associated with burning hydrocarbons for energy. By mandating that certain level of renewable fuel be blended into ethanol, the target of the policy was to reduce carbon dioxide emissions. The failure of this policy to accurately forecast energy prices due to their vast uncertainty has left the policy in a state of rebuilding. The current "reset" period written into the law, effective when mandated levels are inconsistent or fluctuating, calls for a thorough evaluation of the characteristics of industries and the benefits and costs involved in legislation. Renewable Portfolio Standard is a policy similarly designed to produce more renewable electricity and mitigate pollution. Hollingsworth and Rudik (2019) find that a positive impact of the policies on welfare. Further, the German Energiewende is designed to cease nuclear energy consumption, so further research including nuclear energy may be beneficial in examining its own substitutability of various energy forms in a comparable market. However, using a demand system framework in that context may be inappropriate because consumers arguably have less control over their level of nuclear energy usage than electricity, natural gas, petroleum, or biomass.

The EASI demand system is desirable because it allows for nonlinear Engel curves. Lewbel and Pendakur (2009) and Banks et. al (1997) agree that a restrictive specification of this curvature can bias results. Further, the EASI demand system nests the Almost Ideal Demand System of Deaton and Muellbauer (1980) as well as the Quadratic AIDS (QUAIDS) model. Labandeira et al (2017) examine over 1,700 studies on estimating energy elasticities and fail to mention one that employs such a method. Since this recent meta-analysis, Tovar Reaños and Wölfing (2018) study German household demand using the EASI framework and Renner et al. (2018) use the QUAIDS to analyze Mexican household demand. Therefore, this paper makes a direct contribution to a relatively new way of examining energy demand. Woo et al. (2018) recently published a paper using a similar data source (monthly panel from Energy Information Administration over 2000-2016) analyzing the residential, commercial, and industrial sector with a Generalized Leontief (GL) framework. They suggest that a demand system specification is a valuable avenue of future research. The present contribution to the literature is applying an EASI demand system to analyze the United States residential energy demand using EIA data. A feasible extension of this analysis would be to apply the model to similar countries. This paper also contributes methodologically employing the Differential Exact Affine Stone Index (DEASI) demand system. The flexibility of these systems to allow for demographic characteristics and weather make them desirable in the context of analyzing residential energy consumption.

Methodology

Budget shares capture a consumer's decision-making process. The EASI demand system arose out of a need to handle highly non-linear Engel curves, which describes budget shares as a function of expenditure. Tovar Reaños and Wölfing (2018) show that Engel curves for energy goods fit this description.

Conditional demand systems rely on the assumption of separability. This implies that a consumer makes purchasing decisions about groups of individual goods separate from others. In this case, the consumer makes the first stage purchasing decision for all other household goods besides primary energy goods, and then decides how to allocate resources among natural gas, petroleum, and biomass in the second stage. While this approach is empirically tractable it is flawed in that it does not consider all possible goods available to the consumer. Tovar Reaños and Wölfing (2018) suggest two specifications that are common. The first is to aggregate all household goods into the first stage as well as one unique household energy good. The second is to assume separability and have a separate second stage decision for unique primary energy goods. This paper adopts the latter. This restriction cannot be tested statistically and is thus maintained. The adding up restriction on the budget shares is also maintained. However, symmetry and homogeneity are tested and implemented.

The EASI model has been widely utilized recently in the agricultural and resource economics literature. Hovhannisyan and Shanoyan (2018), Hovhannisyan et. al (2019), Boonsaeng and Carpio (2019) and examine food demand in China, Russia, and the United States, respectively. As mentioned before Tovar Reaños and Wölfing (2018) is the only study to knowledge to incorporate the EASI model to energy demand.

The derivation of the EASI system begins by assuming the following cost function like Lewbel and Pednakur (2009) and Pendakur (2009):

$$1)\quad \ln[C(\mathbf{p}, y)] = y + \sum_{i=1}^{I} m_i\,(y, \mathbf{z}) \ln(p_i) + \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J} a_{ij} \ln(p_i) \ln(p_j) + \sum_{i=1}^{I} \varepsilon_i \ln(p_i)$$

where **p** is an I vector of energy prices and y is log expenditures. The y variable can also be interpreted as implicit utility. The **z** term is a vector of demographic characteristics.

To specify y according to the EASI system and the m function, which controls the flexible parametrization of the model:

$$2)\quad \tilde{y} = \ln(x) - \sum_{i=1}^{I} w_i \ln(p_i)$$

$$3)\quad m_i = \sum_{r=0}^{R} b_r \ln\tilde{y}^r + \sum_{l=1}^{L} d_{il} z_l \ln\tilde{y} + \sum_{l=1}^{L} g_{il} z_l$$

where x is total expenditure for all energy goods, $w_i$is the budget share of the ith good, $\tilde{y}^r$is the polynomial of the Stone Price Index-deflated log expenditures, and p_j is the price of the jth good. Additive demographic shifters and weather are applied to the model through z_i. Further, these are interacted with the new expenditure term. Lewbel and Pendakur (2008) find that the name “Exact Affine Stone Index” refers the AIDS family of models invoking the Stone Price Index as an approximate deflator to the expenditure term. In the EASI systems, this relationship is exact. By subtracting $\sum_{i=1}^{I} w_i \ln(p_i)$, $\tilde{y}$ can be interpreted as the natural log of real expenditure.

By applying Shephard’s lemma to (1) and plugging in (2) and (3) the EASI model of Lewbel and Pendakur (2008) can be written as follows:

$$4)\quad w_i = b_0 + \sum_{r=1}^{R} b_{ir} \ln\tilde{y}^r + \sum_{j=1}^{J} a_{ij} \ln p_j + \sum_{l=1}^{L} d_{il} z_l \ln\tilde{y} + \sum_{l=1}^{L} g_{il} z_l + e_i\ ; i, j = 4$$

This paper uses state dummies and the degree Fahrenheit anomaly from the mean temperature for each state every year. There also exists interactions between these variables and the model's expenditure term. Lewbel and Pendakur (2008) and Pendakur (2009) show that the approximate model performs consistently well versus a more complex specification. Alston et al. (1994) and Pendakur (2009) show how to handle potential endogeneity of budget shares in the Stone price index by using an iterated Three Stage Least Squares approach. This paper will adopt a similar approach.

The EASI demand system is subject to theoretical restrictions of symmetry, homogeneity, and Engel aggregation. The symmetry restrictions are:

5) $a_{ij} = a_{ji}$ and $\sum_{k=1}^{I} a_{kj} = \sum_{h=1}^{I} a_{ih} =$ for all $i, j, i \neq j$

The other general restrictions are:

6) $\sum_{i=1}^{I} b_{ir} = 0$ for $r = 1, \dots, R$; and $\sum_{i=1}^{I} b_{i0} = 1$

and

7) $\sum_{i=1}^{I} d_{il} = \sum_{i=1}^{I} g_{il} = 0$ for all i

This paper is only the second, after Tovar Reaños and Wölfing (2018) to implement the EASI model to residential energy demand analysis. Testing of the theoretical restrictions provides a direct extension to Woo et al. (2018), who call for this type of analysis. Therefore, this paper directly fits into the household energy demand literature. The symmetry restrictions are tested, and the null hypothesis of theoretical symmetry is rejected for all pairs of goods with a p-value of <0.001 except for the natural gas and wood combination. There is failure to reject the null hypothesis corresponding to those goods with a p-value of 0.2676. Homogeneity is also rejected with a p-value below 0.05.

Economic theory provides no *ex ante* method of determining a correct model between the EASI and other choices. As mentioned before, the EASI model and the flexibility in specification of the polynomial expenditure term is an advantageous property of the model. Alston and Chalfant (1993) propose a compound model of the Rotterdam and LA-AIDS that utilizes likelihood ratio tests to develop the appropriate model. The EASI demand system nests the Almost Ideal Demand System of Deaton and Muellbauer (1980) as well as the QUAIDS model of Banks et. al (1997). A likelihood ratio test can be used to determine the appropriate model. A likelihood ratio test of the null hypothesis that the restricted EASI is statistically like the unrestricted case is rejected with a p-value of <0.001. A similar test is used to compare the QUAIDS to the restricted EASI model. There is failure to reject the null hypothesis that the QUAIDS and restricted EASI model are statistically similar with a p-value of 0.3718. This result makes sense because the QUAIDS is approximately the EASI model when r=2. Finally, the QUAIDS and the AIDS model are compared via the LR test. The null hypothesis that the two models are similar is rejected with a p-value of <0.001. Given these results, the restricted EASI will be used for coefficient estimation.

Own price elasticities from this EASI system can be calculated as followed:

8) $\eta_{ii} = -1 + \frac{a_{ii}}{w_i} - (b_{i1} + 2b_{i2} + \ldots + Rb_{iR} + \sum_{l=1}^{L} d_{il} z_l)$

Expenditure elasticities are:

9) $A_i = \frac{b_{i1} + 2b_{i2} + \ldots + Rb_{iR} + \sum_{l=1}^{L} d_{il} z_l}{w_i} + 1$

Since the price elasticities from this system are Marshallian, the Hicksian price elasticities can be obtained by using the Slutsky equation:

$$10)\ \eta_{ij}^{*} = \eta_{ij} + w_j A_i$$

Cross-price elasticities are:

$$11)\ \eta_{ij} = \frac{a_{ij} - (b_{i1} + 2b_{i2} + \ldots + Rb_{iR} + \sum_{l=1}^{L} d_{il} z_l) w_j}{w_i}$$

The formulas for the QAIDS model are given by setting $b_{ir} = 0$ for all r > 2. The AIDS model formulas can be derived by further setting $b_{ir} = 0$ for all r > 1.

Data and Descriptive Statistics

Energy data comes via The United States Energy Information Administration for the years 1980 until 2017 for the 48 lower United States on prices and expenditures in the residential sector for natural gas, distillate fuel oil, hydrocarbon gas liquid, and wood. Electricity is not included in this analysis due to its classification by EIA as a secondary energy good. Using this data, variables for budget shares, and Stone (1954) Price Index are calculated. Price is in dollars per million Btu. According to the EIA there is no deflation of the prices. Since the expenditure data is at the aggregate state level, population data for the same time horizon is obtained from the Federal Reserve Economics Database (FRED) to convert it to per capita. The weather data is from the National Oceanic and Atmospheric Administration and includes temperature in degrees Fahrenheit, degrees Fahrenheit abnormality from mean, and ranking of years by warmth. Table 1 provides summary statistics for the data.

No paper to knowledge has implement a differential version of the EASI demand system. A Differential Exact Affine Stone Index (DEASI) demand system is identified to examine the robustness of elasticities. The nearest departures of the baseline model to this paper are

Hovhannisyan and Vardan (2017) and Hovhannisyan and Shanoyan (2019) who use a Generalized Exact Affine Stone Index (GEASI) demand system.

Figure 1 says since the 1980's, the natural gas and hydrocarbon gas liquid shares of budget have increased. Distillate fuel oil shares have declined, while wood shares have remained relatively unchanged.

Figure 2 shows the mean budget shares for the selected forms of energy. Natural gas dominates with an average budget share of 65.11%, followed by hydrocarbon gas liquids at 16.49%, distillate fuel oil at 15.66%, and wood at 2.74%. These figures may be sensitive to the time horizon of the data, as major global events such as the OPEC energy crisis in the 1970's and 1980's, financial crisis of 2008, and regulatory action. Dixit and Pindyck (1998), Postali and Picchetti (2006), and Ghoshray and Johnson (2010) claim that large sample sizes in time-series data (over 100 years) are less likely to yield fruitful insight about the post OPEC energy demand. The DEASI specification will allow for the modelling of a consumer taste or preference trend over time.

Results and Discussion

All estimations were performed using iterated 3 Stage Least Squares in Stata 15. The model is semi-log. Table 2 shows the results from the restricted EASI model with r=3. The own price coefficients for natural gas, distillate fuel oil, and wood are all positive and statistically significant. This makes intuitive sense. As prices go up, the expected budget share for that good increases as well. This is not the case for hydrocarbon gas liquid, whose sign is negative.

The estimates for the DEASI are reported in Tables 3 to analyze the robustness of the coefficients to the model specification. The estimates appear stable to parameterization.

Table 4 is list of model specifications and Marshallian own price elasticity estimates for the energy goods. It says that a one percent increase in the price of natural gas will correspond with a 0.32% decrease in quantity demanded for natural gas. A one percent increase in the price of distillate fuel will correspond with a 0.62% decrease in its own quantity demanded. A one percent increase in the price of hydrocarbon gas liquids will correspond with a 1.3% decrease in its own quantity demanded. A one percent increase in the price of wood will correspond with a 3.4% increase in quantity demanded of wood.

Elasticities in Table 6 appear relatively stable to parameterization of the EASI and DEASI models. Balestra and Nerlove (1966) find natural gas price elasticity to be -.06 for entire country demand. Berndt and Watkins' (1997) also find a similar result for Canadian residential and industrial natural gas price elasticity of -.7. Alberini and Filippini (2011) offer a price elasticity of demand of -.6 for residential sector in the United States. Some estimates may be biased towards zero due to not accounting for exogenous weather. Labandeira et al (2017) find the natural gas own price elasticities to be between -.184 and -.566. Petroleum price elasticity has been examined in numerous meta-analyses; Espey (2008), Graham and Glaister (2002), and Hanly et al. (2002). Brons et al. (2008) examine over 40 studies and find a distribution of elasticities through the literature.

Table 5 is list of model specifications and expenditure elasticity estimates for the energy goods. It says that a one percent increase in the expenditure will correspond with a 0.75% increase in quantity demanded for natural gas. A one percent increase in expenditure will correspond with a 1.6% increase in quantity demanded for hydrocarbon gas liquids. A one percent increase in expenditure will correspond with a 1.61% increase in quantity demanded for distillate fuel oil. A one percent increase in expenditure will correspond with a 0.2% decrease in quantity demanded for wood. All elasticities in Table 5 appear relatively stable to parameterization of the models, like the Marshallian own price elasticities. Burke and Yang (2016) in a meta-analysis find the average natural gas income elasticity in the literature to be greater than 1. Csereklyei et al. (2016) and Burke and Csereklyei (2016) find the natural gas "GDP elasticity" to be .7. While this differs mechanically from the expenditure elasticity, the authors suggest the importance of including weather into the model, further validation of this framework.

Table 6 is list of model specifications and Hicksian own price elasticity estimates for the energy goods. It says that a one percent increase in the price of natural gas will correspond with a 0.16% increase in quantity demanded for natural gas. A one percent increase in the price of hydrocarbon gas liquid will correspond with a 1.01% decrease in its quantity demanded. A one percent increase in the price of distillate fuel will correspond with a 0.37% decrease in quantity demanded for it. A one percent increase in the price of wood will correspond with a 3.4% increase in the quantity demanded for wood. Each elasticity sign make sense according to demand theory except for natural gas and wood. The elasticities in Table 6 appear relatively stable to parameterization.

Table 7 shows elasticity of substitution estimates for the pairs of energy goods for each specification, respectively.

The sign for the distillate fuel/hydrocarbon liquid elasticity is negative, implying that the goods are complements. This makes sense, as both products are derivatives of petroleum. Natural gas is a substitute for each of the energy goods in the system. If regulators were interested in substituting away from petroleum-based products, there is evidence to suggest that a tax on them would lead to consumer substitution towards natural gas. Further research should be conducted into the substitutability of residential energy goods before energy transition policy can be definitively implemented.

Conclusion

Further research into the area of energy economics requires an understanding of the determinants shifting consumer and producer landscape. This paper contributes to the literature by being the first to provide elasticity estimates for energy demand in the United States residential sector using the EASI demand model controlling for weather. Not controlling for weather may bias estimates towards zero. Energy goods at the residential level in the United States are relatively price inelastic.

This paper also contributes methodologically to the literature by being the first to implement the Differential Exact Affine Stone Index (DEASI) demand system. Further research can be done to compare elasticity estimates from this specification to the EASI and GEASI systems.

A similar analysis can be performed on the transportation sector data from the EIA. The transportation sector may be of interest as it is responsible for approximately 27% of the carbon dioxide emissions in the United States. An extension of this paper's analysis to the commercial, industrial, and transportation sectors would fully address the Woo et al. (2018) suggestion to extend their analysis with demand systems.

American policymakers and researchers can use elasticities in Regulatory Impact Analysis for future environmental work. Based on the results, natural gas and wood are both substitutes for petroleum-based hydrocarbon gas liquids and distillate fuel oil. The petroleum-based products are found to be complements, which is not surprising. Due to the inconclusiveness of the literature regarding the complementarity and substitutability of various energy inputs, further research should be undertaken before committing tax dollars to policy implementation.

References

Alberini, Anna, and Massimo Filippini. "Response of residential electricity demand to price: The effect of measurement error." *Energy economics* 33.5 (2011): 889-895.

Alston, Julian M., and James A. Chalfant. "The silence of the lambdas: A test of the almost ideal and Rotterdam models." *American Journal of Agricultural Economics* 75.2 (1993): 304-313.

Alston, Julian M., Kenneth A. Foster, and Richard D. Green. "Estimating elasticities with the linear approximate almost ideal demand system: some Monte Carlo results." The review of Economics and Statistics (1994): 351-356.

Balestra, Pietro, and Marc Nerlove. "Pooling cross section and time series data in the estimation of a dynamic model: The demand for natural gas." *Econometrica: Journal of the Econometric Society* (1966): 585-612.

Banks, James, Richard Blundell, and Arthur Lewbel. "Quadratic Engel curves and consumer demand." *Review of Economics and statistics* 79, no. 4 (1997): 527-539.

Barnett, William A. "Theoretical foundations for the Rotterdam model." *The Review of Economic Studies* 46.1 (1979): 109-130.

Barten, Anton P. "Consumer demand functions under conditions of almost additive preferences." *Econometrica: Journal of the Econometric Society* (1964): 1-38.

Barten, Anton P. "Consumer allocation models: choice of functional form." *Empirical Economics* 18, no. 1 (1993): 129-158.

Berndt, Ernst R., and Gordon Campbell Watkins. *Energy prices and productivity trends in the Canadian manufacturing sector, 1957-76: some exploratory results*. Economic Council of Canada], 1981.

Boonsaeng, Tullaya, and Carlos E. Carpio. "A comparison of food demand estimation from Homescan and Consumer Expenditure Survey data." *Journal of Agricultural and Resource Economics* 44, no. 1835-2019-066 (2019): 117-140.

Brons, Martijn, et al. "A meta-analysis of the price elasticity of gasoline demand. A SUR approach." *Energy Economics* 30.5 (2008): 2105-2122.

Burke, Paul J., and Zsuzsanna Csereklyei. "Understanding the energy-GDP elasticity: A sectoral approach." *Energy Economics* 58 (2016): 199-210.

Burke, Paul J., and Hewen Yang. "The price and income elasticities of natural gas demand: International evidence." *Energy Economics* 59 (2016): 466-474.

Csereklyei, Zsuzsanna, et al. "The effect of economic growth, oil prices, and the benefits of reactor standardization: Duration of nuclear power plant construction revisited." *Energy Policy* 91 (2016): 49-59.

Deaton, Angus, and John Muellbauer. *Economics and consumer behavior*. Cambridge university press, 1980.

Deaton, Angus, and John Muellbauer. "An almost ideal demand system." *The American economic review* 70.3 (1980): 312-326.

Dixit, Avinash K., and Robert S. Pindyck. *Expandability, reversibility, and optimal capacity choice*. No. w6373. National Bureau of Economic Research, 1998.

Espey, Molly. "Gasoline demand revisited: an international meta-analysis of elasticities." *Energy Economics* 20.3 (1998): 273-295.

Fousekis, Panos, and Brian J. Revell. "Meat demand in the UK: A differential approach." *Journal of Agricultural and Applied Economics* 32, no. 1 (2000): 11-19.

Ghoshray, Atanu, and Ben Johnson. "Trends in world energy prices." *Energy Economics* 32.5 (2010): 1147-1156.

Goodwin, P. and Dargay, J. and Hanly, M. (2004) Elasticities of road traffic and fuel consumption with respect to price and income: a review. *Transport Reviews*, 24 (3). pp. 275-292. ISSN 01441647

Graham, D., and S. Glaister. "Review of income and price elasticities of demand for road traffic (Contract number PPAD 9/65/93)." *Centre for Transport Studies, Imperial College of Science, Technology and Medicine, Imperial College, Final Report* 18 (2002).

Hollingsworth, Alex, and Ivan Rudik. "External impacts of local energy policy: The case of renewable portfolio standards." *Journal of the Association of Environmental and Resource Economists* 6, no. 1 (2019): 187-213.

Hovhannisyan, Vardges, and Aleksan Shanoyan. "Addressing Pre-Commitment Bias with a Generalized EASI Model: An Application to Food Demand in Russia." *Journal of Agricultural and Resource Economics* 44, no. 1835-2019-064 (2019): 80-97.

Hovhannisyan, Vardges, Sachintha Mendis, and Chris Bastian. "An econometric analysis of demand for food quantity and quality in urban China." *Agricultural economics* 50, no. 1 (2019): 3-13.

Hovhannisyan, Vardges, and Vardan Urutyan. "A Structural Model of Pre-committed Demand: The Case of Food Demand in China." (2017).

https://www.eia.gov/energyexplained/natural-gas/use-of-natural-gas.php

https://www.eia.gov/energyexplained/oil-and-petroleum-products/use-of-oil.php

https://www.eia.gov/energyexplained/renewable-sources/

https://www.eia.gov/energyexplained/us-energy-facts/

https://www.eia.gov/energyexplained/use-of-energy/homes.php

Huntington, Hillard G., James J. Barrios, and Vipin Arora. "Review of key international demand elasticities for major industrializing economies." *Energy Policy* 133 (2019): 110878

Labandeira, Xavier, José M. Labeaga, and Xiral López-Otero. "A meta-analysis on the price elasticity of energy demand." *Energy Policy* 102 (2017): 549-568.

Lade, Gabriel E., C. Y. Cynthia Lin Lawell, and Aaron Smith. "Designing climate policy: Lessons from the Renewable Fuel Standard and the blend wall." *American Journal of Agricultural Economics* 100.2 (2018): 585-599.

Lee, Chien-Chiang, and Yi-Bin Chiu. "Nuclear energy consumption, oil prices, and economic growth: Evidence from highly industrialized countries." *Energy Economics* 33, no. 2 (2011): 236-248.

Lewbel, Arthur, and Krishna Pendakur. "Tricks with Hicks: The EASI demand system." *American Economic Review* 99, no. 3 (2009): 827-63.

Lewbel, Arthur, and Krishna Pendakur. "Tricks with Hicks: The EASI implicit Marshallian demand system for unobserved heterogeneity and flexible Engel curves." *American Economic Review, forthcoming* (2008).

Keller, Wouter Jacques, and Jan Van Driel. "Differential consumer demand systems." *European Economic Review* 27, no. 3 (1985): 375-390.

Maynard, Leigh J., and Venkat N. Veeramani. "Price sensitivities for US frozen dairy products." *Journal of Agricultural and Applied Economics* 35, no. 3 (2003): 599-609.

Mountain, Dean, and Cheng Hsiao. "A combined structural and flexible functional approach for modeling energy substitution." *Journal of the American Statistical Association* 84.405 (1989): 76-87.

Neves, Pedro Duarte. "A class of differential demand systems." *Economics Letters* 44, no. 1-2 (1994): 83-86.

Pendakur, Krishna. "Chapter 7 EASI Made Easier." In *Quantifying Consumer Preferences,* pp. 179-206. Emerald Group Publishing Limited, 2009.

Postali, Fernando AS, and Paulo Picchetti. "Geometric Brownian motion and structural breaks in oil prices: a quantitative analysis." *Energy Economics* 28.4 (2006): 506-522.

Renner, Sebastian, Jann Lay, and Hannes Greve. "Household welfare and CO2 emission impacts of energy and carbon taxes in Mexico." *Energy Economics* 72 (2018): 222-235.

Schmalensee, Richard, and Robert Stavins. *Lessons learned from three decades of experience with cap-and-trade*. No. w21742. National Bureau of Economic Research, 2015.

Stone, Richard. "Linear expenditure systems and demand analysis: an application to the pattern of British demand." *The Economic Journal* 64.255 (1954): 511-527.

Theil, Henri. "The information approach to demand analysis." *Econometrica: Journal of the Econometric Society* (1965): 67-87.

Tovar Reaños, M.A.T. and Wölfing, N.M., 2018. Household energy prices and inequality: Evidence from German microdata based on the EASI demand system. *Energy Economics*, *70,* pp.84-97.

Woo, Chi Keung, et al. "Price elasticities of retail energy demands in the United States: New evidence from a panel of monthly data for 2001–2016." *Applied energy* 222 (2018): 460-474.

Appendix 1: Tables

Table 1: Descriptive Statistics

Variable	Obs	Mean	Std.Dev.	Min	Max
NG Share	1824	.651	.21	.012	.97
HGL Share	1824	.165	.103	.009	.581
DFO Share	1824	.157	.206	0	.917
Wood Share	1824	.027	.028	.001	.256
NG P	1824	8.328	3.447	2.38	20.88
Wood P	1824	5.799	3.48	1.75	17.11
DFO P	1824	11.909	7.243	0	30.04
HGL P	1824	15.975	8.181	2.51	42.51
NG PCX	1824	321.845	514.087	2.669	5657.646
DFO PCX	1824	88.004	212.557	0	3346.12
HGL PCX	1824	65.837	104.15	.375	949.109
Wood PCX	1824	11.252	22.577	.085	302.167
Temp anom.	1824	.265	1.307	-4.3	4.2
Easi Y	1824	4.746	1.338	1.438	8.483

Note: Price is \$/million Btu, Per capital expenditure (PCX) is \$, Temp anom. is degrees Fahrenheit anomaly

Table 2: Results from Restricted EASI from Equation 2, r=3

	NG	HGL	DFO	Wood
ln p NG	0.310***	-0.116***	-0.0712**	-0.123***
ln p HGL	-0.116***	-0.0776*	0.0962***	0.0975***
ln p DFO	-0.0712**	0.0962***	0.0685*	-0.0935***
ln p Wood	-0.123***	0.0975***	-0.0935***	0.119***
Y	0.00186	0.0905**	-0.0891*	-0.0032
Y2	0.0123	-0.0260***	0.0124	0.0013
Y3	-0.00131	0.00189***	-0.000452	-0.000131
temp from mean	-0.00410	0.0126***	- 0.00827***	-0.00021
Constant	0.754***	0.0628	0.160	0.0234
AIC	-15394.7			
BIC	-15130.4			

Note: FERC dummies not displayed but included in model.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Results from DEASI

	NG	HGL	DFO	Wood
D ln p NG	0.505***	-0.315***	-0.138***	-0.0515***
D ln p HGL	-0.369***	0.407***	-0.0258	-0.0129*
D ln p DFO	-0.107***	-0.0661***	0.179***	-0.00524
D ln p Wood	-0.0812**	-0.0131	-0.00261	0.0969***
D Y	-0.143**	0.0633	0.0907*	-0.0105
D Y2	0.0194	-0.00426	-0.0170*	0.00183
D Y3	-0.00113	-0.0000112	0.00126*	-0.000125
D temp from mean	-0.00553*	0.00326	0.000675	0.00159**
Constant	0.00114	0.00307**	-0.00441***	0.00199
AIC	-25187.8			
BIC	-25023.4			
R-squared	0.375	0.242	0.184	0.116

Note: FERC dummies not included due to time-invariance.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Marshallian Own-Price Elasticities by Specification

	Coef.	Std.Err.	P-value	[95%Conf.	Interval]
<u>EASI Model</u>					
EtaNG	-0.32	0.26	0.209	-0.82	-0.18
EtaH	-1.27	0.28	0.000	-1.82	-0.72
EtaD	-0.62	0.28	0.025	-1.16	-0.08
EtaW	3.42	0.28	0.000	2.87	3.97
<u>DEASI Model</u>					
EtaNG	-0.12	0.06	0.060	-0.24	0.05
EtaH	1.41	0.13	0.000	1.15	1.68
EtaD	0.08	0.09	0.367	-0.09	0.25
EtaW	2.60	0.25	0.000	2.10	3.09

Table 5: Expenditure Elasticities by Specification

	Coef.	Std.Err.	P-value	[95%Conf.	Interval]
EASI Model					
ANG	0.75	0.39	0.055	-0.02	1.51
AH	1.60	0.99	0.106	-0.34	3.54
AD	1.61	1.24	0.194	-0.82	4.05
AW	-0.20	1.71	0.907	-3.56	3.16
DEASI Model					
ANG	0.94	0.07	0.000	0.80	1.09
AH	2.14	0.23	0.000	1.68	2.59
AD	0.75	0.18	0.000	0.390	1.11
AW	3.86	0.40	0.000	3.05	4.67

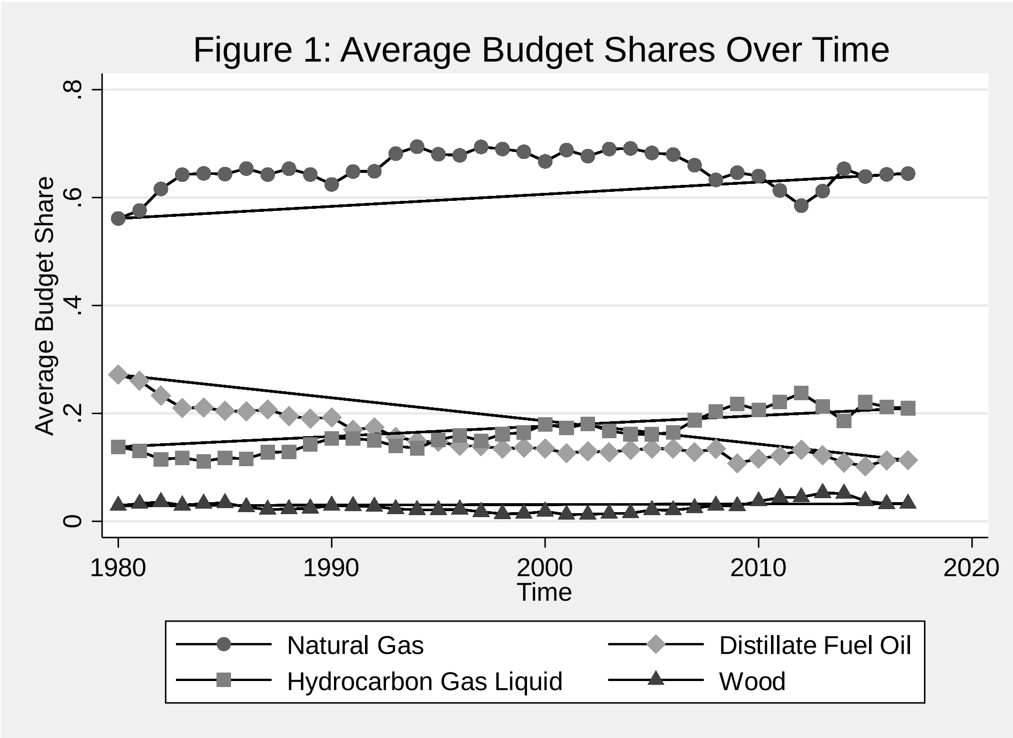
Table 6: Hicksian Own-Price Elasticities by Specification

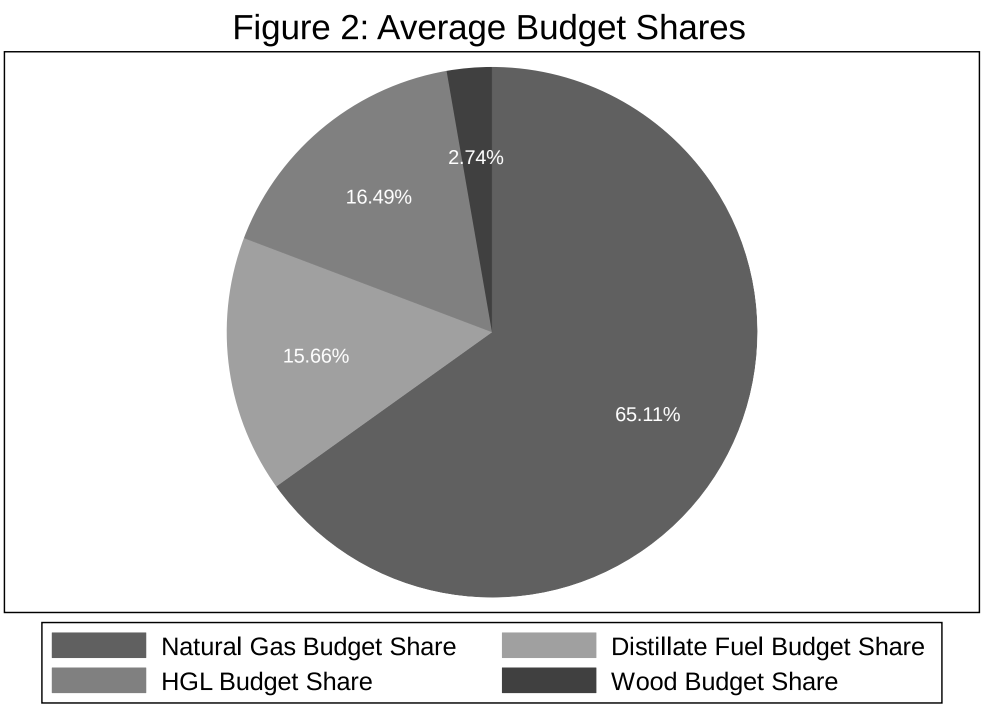
	Coef.	Std.Err.	P-value	[95%Conf.	Interval]
EASI Model					
EtaStarNG	0.16	0.06	0.005	0.05	0.28
EtaStarH	-1.01	0.23	0.000	-1.46	-0.56
EtaStarD	-0.37	0.18	0.044	-0.72	-0.01
EtaStarW	3.42	0.28	0.000	2.88	3.96
DEASI Model					
EtaStarNG	-0.68	0.06	0.000	-0.80	-0.56
EtaStarH	0.69	0.13	0.000	0.43	0.95
EtaStarD	-0.64	0.09	0.000	-0.81	-0.47
EtaStarW	1.60	0.25	0.000	1.10	2.09

Table 7: Allen Substitution Elasticities by Specification

	Coef.	Std.Err.	P-value	[95%Conf.	Interval]
EASI Model					
SigmaNGH	1.00	0.35	0.004	0.32	1.69
SigmaNGD	1.06	0.37	0.004	0.33	1.78
SigmaNGW	6.14	2.14	0.004	1.94	10.34
SigmaHW	-37.34	8.56	0.000	-54.11	-20.57
SigmaHD	-6.42	1.47	0.000	-9.31	-3.54
SigmaDW	-13.60	6.74	0.044	-26.81	-0.38
DEASI Model					
SigmaNGH	2.07	0.21	0.000	1.66	2.47
SigmaNGD	2.23	0.21	0.000	1.81	2.66
SigmaNGW	18.89	1.23	0.000	16.47	21.30
SigmaHW	25.49	4.92	0.000	15.85	35.14
SigmaHD	4.38	0.85	0.000	2.73	6.04
SigmaDW	-23.82	3.19	0.000	-30.07	-17.57

Appendix 2: Figures





Chapter 3: Air Pollution, Investor Sentiment and Excessive Returns

Abstract

This paper extends the asset pricing literature by offering a proprietary index of negative investor sentiment linked to carbon monoxide (CO), nitrogen dioxide (NO2), ozone particle (O3), 2.5 mm particulate matter (PM2.5), and sulfur dioxide (SO2) levels; determining the link between New York City air pollution and stock market returns. Food products and wholesale portfolio returns on average increase with enhanced negative investor sentiment. This is consistent with behaviors associated with psychological stress, like binge eating and shopping sprees. Personal services portfolio returns decrease on average with increased negative investor sentiment, consistent with behavioral isolationism.

Keywords

Asset pricing models, investor sentiment, air pollution

Introduction

How does air pollution affect the stock market? The objective of this paper is to assess the relationship between pollution, investor sentiment, and stock market returns. The United States stock market is the holding choice of over $30 trillion in wealth. A risk averse investor responds to uncertainty by his or her willingness to pay for a risk premium to achieve a more certain state. This is the fundamental underlying of the Capital Asset Pricing Model (CAPM) of Sharpe (1963). Their model says that the expected return of a risky asset can be explained by a composition of the difference between returns and risk-free rate of return plus the risk premium. Fama and French (1993) extend this relationship to include the difference between returns of portfolios diversified with small stocks and big stocks respectively and the difference between returns on high book-to-market value stocks versus low. Fama and French (2015) continue the extension by adding variables for the difference in return between highly profitable and the least profitable as well as one for the difference in returns for firms who invest aggressively versus conservatively. Investor sentiment may alter asset prices away from their fundamental value if only considering the predefined characteristics. This paper contributes a unique set of negative investor sentiment components to the asset pricing literature, generated using a principle-component analysis of New York City's daily carbon monoxide (CO), nitrogen dioxide (NO2), ozone particle (O3), 2.5 mm particulate matter (PM2.5), and sulfur dioxide (SO2) levels. Using pollution measures instead of weather variations are desirable because it is more feasible to implement a policy of emissions reduction than to control Mother Nature, and thus this paper provides an interesting alternative channel to explore in the welfare implications of emission reduction policies. The new model adds this index to the Fama and French (2015) five factors and is implemented on the 49 industry portfolios provided by Dr. French's website in order to examine if it can adequately explain returns across a wide variety of sectors and securities. The results indicate that sentiment based on pollution can further explain stock market returns and may be useful to implement into a trading strategy. The rest of the paper is organized as follows: a review of the relevant literature relating to stock market return factors, a theoretical derivation of the original CAPM of Sharpe (1963) and extension to serve the purpose of the paper, an overview of the data and empirical methodology, results, discussion, and conclusion.

Review of Literature

Li and Zhang (2019) show a positive impact of air pollution on the disposition effect, which is a behavioral anomaly where traders hold onto assets whose prices are dropping and sell those who are increasing. They show that the effect is larger when measuring air pollution by 10 or 2.5 mm inhalable particulate matter (PM10 or PM2.5), further validating the use of PM25 data in this study. Levy and Yagil (2011), Lepori (2016), Li and Peng (2016), and An et al. (2018) also study the effects of air pollution on the stock market and find negative relationship between air pollution and returns in the United States, Italy, and China. Heyes et al. (2016) examine primarily PM2.5 and finds a robust negative statistically significant relationship. The China AQI in Li and Peng (2016) contains information for carbon monoxide, nitrogen dioxide, ozone particle, 2.5 mm particulate matter, and sulfur dioxide. Their study highlights these as five associated with negative human health consequences, particularly PM2.5 that can infiltrate alveoli and obstruct gas exchange.

Saunders (1993) examines the relationship between sunny days, investor sentiment and expected returns between 1983 and 1989 and finds no significant “sunshine effect.” Hirshleifer and Shumway (2003) found that trading decisions made incorporating sunshine information can increase a portfolio’s Sharpe ratio, measuring excess returns for a given unit of risk (measured in standard deviation of returns). The Britten-Jones test can be used to identify this relationship and is conducted by regression of 1’s on the vector of portfolio returns. They do mention however that these results are sensitive to the frequency of trades if non-trivial transaction costs exist. Their study examines daily market returns in 26 countries from 1982 to 1997. Their use of sunshine in the city where the most active financial exchange exists motivates our use of using New York City weather data. Chang et al. (2008) conclude that increased cloud cover in New York City is associated with increased stock volatility. Their results confirm those of Saunders (1993) and Hirshleifer and Shumway (2003) in that there are linkages between daily weather patterns and asset returns. Trombley (1997) critiques Saunders (1993) paper by saying the distribution of cloudiness in them lends itself to statistically significant results by comparing

20% cloudiness to 100%. We address this concern by using widely dispersed pollution data. Cao and Wei (2001) examine temperature effects on stock market returns while remaining consistent with the trend in the literature to include data from the major market city. Kliger and Levy (2003) find that increased cloud cover is related to increased investor perceived probabilities of negative events. They find that higher temperatures are associated with apathy and lower returns while adjacently lower temperatures imply aggression and more risk seeking and higher returns. Dowling and Lucey (2008) and Dowling and Lucey (2008) study wind, precipitation, and geomagnetic storms and find that they are relevant drivers of increased volatility of individual indices. Loughran and Schultz (2004) find that blizzards and cloudy days in New York are associated with marginally lower stock returns. One interesting question arises when considering cloud covers studies. How much of the cloud cover can be attributed to air pollution? While the literature appears deeper regarding the effects of weather rather than pollution on the stock market, the detrimental physical health effects of pollution in addition to the adverse mental health effects from clouds that could also be from pollution signal an importance for the field to continue to understand these relationships with the goal to improve human welfare across the multiple aforementioned avenues.

Jaffe, Westerfield, and Ma (1989); Wang, Li, and Erickson (1997); and Pettengill (2003) examine what is referred to as the Monday effect, which originally said there is a decline in labor productivity on this day relative to others. It has been shown to have the reversal effect. Schultz (1985), Ariel (1987), Kramer (2000) describes another seasonal effect pervasive in the literature, the January effect. This effect says that unusually high returns are observed in January relative to other months.

Other examples of seasonal occurrences that may explain returns are Kamstra, Kramer, and Levy (2000) and Kamstra, Kramer, and Levi (2003) who find that daylight savings time and thus shortened days lead to seasonal affective disorder (SAD), associated with depression, which causes an increase in risk aversion. This heightened risk aversion thus leads to increased variability or volatility in asset returns. Dowling and Lucey (2008) and Dowling and Lucey (2008) also examine daylights saving time and lunar phases with similar results. Loughran and Schultz (2004) in addition to their weather results find that trading is slowed in cities with high Jewish populations on Yom Kippur. Also, Dichey and Janes (2003); Yuan et. al (2001); and Keef and Khaled (2011) among others study lunar phases of the moon and stock market returns and provide evidence for moon effects. These studies support the findings that returns are higher on new moon days but argue that it is difficult to imagine that, the Monday effect, or the turn-of-the-month effect as significant drivers off inefficient markets. I agree with this sentiment and would argue that as more recent technological advances such as algorithmic trading strategies, near-zero cost investment platforms and free financial literacy training mobile applications continue to progress, the increased financial savvy of investors across investors of all skill levels will increasingly diminish these particular inefficiencies and thus do not consider these variables in the paper.

Theoretical Framework

The theoretical model relies on fundamental equation of the Sharpe (1963) CAPM:

1) $E(r_j) = a_j + r + [E(r_m) - r]\beta_j, for\ each\ j.$

The standard assumption is that if the model adequately explains returns then alpha should be zero. The arguments against the efficient market hypothesis can be found in Lee et al. (1991) and

Barberis et al. (2005). They find that asset returns are indeed influenced by non-common fundamental risk. These deviations from the assumption provide the Fama and French (1992) framework and allow for further analysis.

The Three Factor Model of Fama and French (1993) includes size (SMB)and book-to-market value (HML) factors to the original CAPM specification. They find significant explanatory power of these factors towards excess return, stating the book-to-market value is positively correlated with asset returns. The same holds true for firm size as measured by market equity.

2) $E(r_j) - r = a_j + [E(r_m) - r]\beta_{j1} + SMB\beta_{j2} + HML\beta_{j3},$ for each j.

Fama and French (2016) extend the model further to the Five Factor Model to include profitability (RMW) and an investment aggressiveness (CMA) factor. They also note that when implementing this model, the value (book-to-market/HML) may become redundant. As an additional avenue of analysis, this paper explores the issue further.

3) $E(r_j) - r = a_j + [E(r_m) - r]\beta_{j1} + SMB\beta_{j2} + HML\beta_{j3} + RMW\beta_{j4} + CMA\beta_{j5},$ for each j.

By asserting that investor sentiment may further explain excess asset returns, this paper extends the theoretical model to include a sixth factor, a unique measure of investor sentiment (SEN).

4) $E(r_j) - r = a_j + [E(r_m) - r]\beta_{j1} + SMB\beta_{j2} + HML\beta_{j3} + RMW\beta_{j4} + CMA\beta_{j5} + SEN\beta_{j6},$ for each j.

The parameters of the model can be estimated using a regression like Fama and McBeth (1973).

Data

Daily air quality index (AQI) data for carbon monoxide, nitrogen dioxide, ozone particle, 2.5 mm particulate matter, and sulfur dioxide in New York City were obtained from the Environmental Protection Agency on trading days from January 4, 2013 until May 7, 2019. Data across a similar time horizon on portfolio returns and the Fama and French (2016) factors are obtained from Dr. French's personal website. The number of trading days in this sample is 1,617. The period before 2013 is excluded from the analysis to avoid overlap into the financial crisis.

Table 1 in the appendix describes the summary statistics of the dataset. The portfolios were constructed based on the stock's industry SIC code. The portfolio types are; agriculture, food products, candy and soda, beer and liquor, tobacco products, recreation, entertainment, printing and publishing, consumer goods, apparel, healthcare, medical equipment, pharmaceutical drugs, chemicals, rubber and plastic products, textiles, construction materials, construction, steel works, fabricated products, machinery, electrical equipment, automobiles and trucks, aircraft, shipbuilding and railroad equipment, defense, precious metals, non-metallic and industrial mining metal, coal, petroleum and natural gas, utilities, communication, personal services, business services, computer hardware, computer software, electronic equipment, measuring and control equipment, business supplies, shipping containers, transportation, wholesale, retail, restaurants and hotels/motels, banking, insurance, real estate, trading, and other. Figures 1 through 5 show the normalized AQI values of various pollutants over time. Visually there is sparsity in the data and alleviates the concern pressed in Trombley (1997).

Empirical Framework

Principle component analysis is used to create the sentiment components. A Bartlett test of sphericity with a null that the correlation matrix for the given variables is not an identity matrix is rejected. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy is .643 which is greater than .5. The diagonal on the anti-image correlation coefficient matrix can be interpreted as a measuring of sampling adequacy. The values for this diagonal are .66, .6989, .8321, .6487, and .8113 respectively. Further, low values along the diagonal of the residual correlation matrix, a measure of performance of the components in explaining the variation in the original data. These values are all under .0005. These results taken in conjunction imply that a principle component analysis of the data may be appropriate.

The first two principle components have eigenvalues above 1 and represent the points to the left of the "bow" of the scree plot (Figure 6), which are two useful criteria in deciding the number of components to use. These components explain roughly 67 percent of the variation in the initial data and may be adequate for analysis (Table 2).

This technique is used in the asset pricing literature to develop indices of investor sentiment by Baker and Wurgler (2006); Lin, Wang and Cai (2012); Ait-Sahalia and Xiu (2017); Dhaoui an Bensalah (2017); and Gerber et al. (2019). Liew and Budavari (2017) use a combination of proprietary StockTwits data to construct their index and provide another unique example of how to measure sentiment.

The loading plot (Figure 7) and loading table (Table 3) show the makeup of the individual components. Component 1 is as followed (recalling that the input data has been normalized):

5) $\mathrm{COMP1} = .5145\mathrm{CO} + .5207\mathrm{NO2} + .0917\mathrm{OZONE} + .5256\mathrm{PM2.5} + .4237\mathrm{SO2}$

Given the theoretical model and the derivation of principle components, the final model to be estimated is:

6) $E(r_j) - r = a_j + [E(r_m) - r]\beta_{j1} + SMB\beta_{j2} + HML\beta_{j3} + RMW\beta_{j4} + CMA\beta_{j5} + COMP1\beta_{j6}$, for each j.

Results and Discussion

The analysis was performed using Stata 15. The estimated output tables can be found in Table 4.

The results indicate that the investor sentiment index composition in this paper do indeed help explain stock market returns. The 49 industry portfolios examined all display an increase in the R-square value because of the implementation. Further, the F-test for the null hypothesis that the factors are jointly insignificant is rejected. The portfolios where the first principle component is individually statistically significant are food products (positive effect), personal services (negative effect), and wholesale (positive). These portfolios correspond to actions in psychological and health literature taken in companionship with stress. The psychological link to binge eating, isolationism, and shopping sprees are well documented. see Smith et al. (1998); Sanders et al. (2000), and Krueger (1998) respectively. It is also of interest that statistically significant positive alpha was generated for guns, business services, and insurance portfolios.

Conclusion

These results add to the expanding literature on the effects of pollution on stock market returns, specifically providing a link between New York City pollution levels and excessive returns. A unique measure of negative investor sentiment, generated by using a principle-component analysis of New York City's daily carbon monoxide, nitrogen dioxide, ozone, 2.5 mm particulate matter, and sulfur dioxide levels is also contributed. Further research into this subfield may include adding similar pollution variables or investor sentiment data to the index. Finally, this paper gives supporting evidence to emission reduction policies, giving an alternative vehicle to welfare improvement because of successful implementation.

References

Ait-Sahalia, Yacine, and Dacheng Xiu. "Using principal component analysis to estimate a high dimensional factor model with high-frequency data." *Journal of Econometrics* 201, no. 2 (2017): 384-399.

An, N., Wang, B., Pan, P., Guo, K. and Sun, Y., 2018. Study on the influence mechanism of air quality on stock market yield and Volatility: Empirical test from China based on GARCH model. *Finance Research Letters*, *26*, pp.119-125.

Baker, Malcolm, and Jeffrey Wurgler. "Investor sentiment and the cross-section of stock returns." *The journal of Finance* 61, no. 4 (2006): 1645-1680.

Cao, M. and Wei, J., 2005. Stock market returns: A note on temperature anomaly. *Journal of Banking & Finance*, *29*(6), pp.1559-1573.

Chang, S.C., Chen, S.S., Chou, R.K. and Lin, Y.H., 2008. Weather and intraday patterns in stock returns and trading activity. *Journal of Banking & Finance*, *32*(9), pp.1754-1766.

Dhaoui, Abderrazak, and Nesrine Bensalah. "Asset valuation impact of investor sentiment: A revised Fama–French five-factor model." *Journal of Asset Management* 18, no. 1 (2017): 16-28.

Dichev, Ilia D., and Troy D. Janes. "Lunar cycle effects in stock returns." *The Journal of Private Equity* 6, no. 4 (2003): 8-29.

Dowling, Michael, and Brian M. Lucey. "Robust global mood influences in equity pricing." *Journal of Multinational Financial Management* 18, no. 2 (2008): 145-164.

Dowling, Michael, and Brian M. Lucey. "Mood and UK equity pricing." *Applied Financial Economics Letters* 4, no. 4 (2008): 233-240.

Fama, Eugene F., and Kenneth R. French. "The cross-section of expected stock returns." *the Journal of Finance* 47, no. 2 (1992): 427-465.

Fama, Eugene F., and Kenneth R. French. "Common risk factors in the returns on stocks and bonds." *Journal of financial economics* 33, no. 1 (1993): 3-56.

Fama, Eugene F., and Kenneth R. French. "A five-factor asset pricing model." *Journal of financial economics* 116, no. 1 (2015): 1-22.

Fama, Eugene F., and Kenneth R. French. "Choosing factors." *Journal of Financial Economics* 128, no. 2 (2018): 234-252.

Fama, Eugene F., and James D. MacBeth. "Risk, return, and equilibrium: Empirical tests." *Journal of political economy* 81, no. 3 (1973): 607-636.

Gerber, Sander, Babak Javid, Harry Markowitz, Paul Sargen, and David Starer. "A Non-Normal Principal Components Model for Security Returns." *Available at SSRN 3352062* (2019).

Goetzmann, W.N. and Zhu, N., 2005. Rain or shine: where is the weather effect?. *European Financial Management*, *11*(5), pp.559-578.

Heyes, Anthony, Matthew Neidell, and Soodeh Saberian. *The effect of air pollution on investor behavior: Evidence from the S&P 500*. No. w22753. National Bureau of Economic Research, 2016.

Hirshleifer, D. and Shumway, T., 2003. Good day sunshine: Stock returns and the weather. *The Journal of Finance*, *58*(3), pp.1009-1032.

Jaffe, Jeffrey F., Randolph Westerfield, and Christopher Ma. "A twist on the Monday effect in stock prices: Evidence from the US and foreign stock markets." *Journal of Banking & Finance* 13, no. 4-5 (1989): 641-650.

Kamstra, Mark J., Lisa A. Kramer, and Maurice D. Levi. "Losing sleep at the market: The daylight-saving anomaly." *American Economic Review* 90, no. 4 (2000): 1005-1011.

Kamstra, M.J., Kramer, L.A. and Levi, M.D., 2003. Winter blues: A SAD stock market cycle. *American Economic Review*, *93*(1), pp.324-343.

Keef, Stephen P., and Mohammed S. Khaled. "Are investors moonstruck? Further international evidence on lunar phases and stock returns." *Journal of Empirical Finance* 18, no. 1 (2011): 56-63.

Kliger, Doron, and Ori Levy. "Mood-induced variation in risk preferences." *Journal of economic behavior & organization* 52, no. 4 (2003): 573-584.

Krueger, David W. "On compulsive shopping and spending: A psychodynamic inquiry." *American journal of psychotherapy* 42, no. 4 (1988): 574-584.

Lanfear, M.G., Lioui, A. and Siebert, M.G., 2018. Market anomalies and disaster risk: Evidence from extreme weather events. *Journal of Financial Markets*.

Lepori, Gabriele M. "Air pollution and stock returns: Evidence from a natural experiment." *Journal of Empirical Finance* 35 (2016): 25-42.

Levy, Tamir, and Joseph Yagil. "Air pollution and stock returns in the US." *Journal of Economic Psychology* 32, no. 3 (2011): 374-383.

Li, Jennifer, Massimo Massa, Hong Zhang, and Jian Zhang. "Chang Air Pollution, Behavioral Bias, and the Disposition Effect in China." *Journal of Financial Economics* (2019).

Liew, Jim, and Tamas Budavari. "The “Sixth” Factor—A Social Media Factor Derived Directly from Tweet Sentiments." *The Journal of Portfolio Management* 43, no. 3 (2017): 102-111.

Lin, Jianhao, Meijin Wang, and Lingfeng Cai. "Are the Fama–French factors good proxies for latent risk factors? Evidence from the data of SHSE in China." *Economics Letters* 116, no. 2 (2012): 265-268.

Loughran, T. and Schultz, P., 2004. Weather, stock returns, and the impact of localized trading behavior. *Journal of Financial and Quantitative Analysis*, *39*(2), pp.343-364.

Pettengill, Glenn N. "A survey of the Monday effect literature." *Quarterly Journal of Business and Economics* 42, no. 3/4 (2003): 3-27.

Sanders, Christopher E., Tiffany M. Field, Miguel Diego, and Michele Kaplan. "The relationship of Internet use to depression and social isolation among adolescents." *Adolescence* 35, no. 138 (2000): 237-237.

Saunders, E.M., 1993. Stock prices and Wall Street weather. *The American Economic Review*, *83*(5), pp.1337-1345.

Smith, Delia E., Marsha D. Marcus, Cora E. Lewis, Marian Fitzgibbon, and Pamela Schreiner. "Prevalence of binge eating disorder, obesity, and depression in a biracial cohort of young adults." *Annals of Behavioral Medicine* 20, no. 3 (1998): 227-232.

Symeonidis, L., Daskalakis, G. and Markellos, R.N., 2010. Does the weather affect stock market volatility?. *Finance Research Letters*, *7*(4), pp.214-223.

Wang, Ko, Yuming Li, and John Erickson. "A new look at the Monday effect." *The Journal of Finance* 52, no. 5 (1997): 2171-2186.

Yuan, Kathy, Lu Zheng, and Qiaoqiao Zhu. "Are investors moonstruck? Lunar phases and stock returns." *Journal of Empirical Finance* 13, no. 1 (2006): 1-23.

Appendix 1: Tables

Table 1: Descriptive Statistics

Variable	Obs	Mean	Std.Dev.	Min	Max
coaqi	1617	8.772	3.603	2	38
no2aqi	1617	42.244	13.714	15	131
ozoneaqi	1617	47.952	25.719	11	210
pm25aqi	1617	50.838	15.93	16	141
so2aqi	1617	4.989	5.056	0	59
mrf	1617	.051	.832	-4.03	5.06
smb	1617	-.007	.493	-1.63	2.52
hml	1617	-.014	.494	-1.69	2.38
rmw	1617	.004	.324	-1.58	1.63
cma	1617	-.009	.316	-1.32	1.96
rf	1617	.002	.003	0	.01
agric	1617	.022	1.122	-6.43	7.65
food	1617	.038	.882	-5.11	4.73
soda	1617	.043	.882	-7.3	5.29
beer	1617	.058	.825	-4.28	3.04
smoke	1617	.038	1.056	-11.46	4.91
toys	1617	.04	1.391	-8.15	8.22
fun	1617	.088	1.464	-6.75	7.15
books	1617	.024	1.128	-8.72	6.91
hshld	1617	.043	.806	-3.98	4.55
clths	1617	.057	1.156	-6.3	6.36
hlth	1617	.045	1.147	-8.86	4.78
medeq	1617	.076	.972	-4.45	4.92
drugs	1617	.054	1.044	-4.68	6.25
chems	1617	.04	1.061	-4.75	5.37
rubbr	1617	.053	.987	-5.11	3.81
txtls	1617	.048	1.409	-18.31	6.56
bldmt	1617	.044	1.129	-5.21	4.17
cnstr	1617	.039	1.307	-6.2	5.13
steel	1617	.019	1.586	-6.85	8.72
fabpr	1617	.035	1.743	-15.45	9.78
mach	1617	.045	1.143	-5.93	5.23
elceq	1617	.033	1.104	-4.82	4.95
autos	1617	.032	1.219	-5.98	5.29
aero	1617	.075	1.065	-5.36	5.12
ships	1617	.063	1.392	-5.64	7.99
guns	1617	.097	1.042	-5.66	6.22
gold	1617	.003	2.405	-11.76	10.42
mines	1617	.006	1.642	-7.52	10.04
coal	1617	-.055	2.799	-18.44	18.08
oil	1617	.006	1.326	-7.47	6.71
util	1617	.044	.835	-4.44	2.89
telcm	1617	.045	.851	-4.42	3.6
persv	1617	.045	1.091	-4.47	4.35
bussv	1617	.075	.952	-4.4	5.17
hardw	1617	.049	1.169	-7	5.11
softw	1617	.074	1.092	-4.8	6.33
chips	1617	.081	1.17	-7.25	5.91
labeq	1617	.076	1.048	-4.56	5.02
paper	1617	.044	.925	-7.35	4.15
boxes	1617	.05	1.09	-5.55	4.4

trans	1617	.059	1.083	-4.66	5.61
whlsl	1617	.041	.894	-4.18	3.85
rtail	1617	.063	.93	-4.1	6.88
meals	1617	.065	.837	-4.31	3.84
banks	1617	.059	1.096	-6.17	5.07
insur	1617	.071	.88	-4.37	4.26
rlest	1617	.025	1.16	-7.25	6.13
fin	1617	.06	1.158	-6.92	5.23
other	1617	.036	.887	-5.25	4.65
zco	1617	0	1	-1.879	8.112
zno2	1617	0	1	-1.987	6.472
zozone	1617	0	1	-1.437	6.301
zpm25	1617	0	1	-2.187	5.66
zso2	1617	0	1	-.987	10.683
pc1	1617	0	1.466	-3.377	8.737
pc2	1617	0	1.083	-4.123	6.281
pc3	1617	0	.841	-4.067	7.986
pc4	1617	0	.74	-3.124	5.331
pc5	1617	0	.65	-2.022	5.19

Table 2: Proportion of Variance Explained by Components

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	2.148	0.975	0.430	0.430
Comp2	1.174	0.466	0.235	0.664
Comp3	0.708	0.161	0.142	0.806
Comp4	0.547	0.124	0.110	0.915
Comp5	0.423	.	0.085	1.000

Table 3: Loading Table

Variable	Comp1	Comp2	Comp3	Comp4	Comp5	Unexplained
zco	0.514	-0.264	-0.436	-0.337	0.602	0
zno2	0.521	0.098	-0.362	0.721	-0.261	0
zozone	0.092	0.858	0.176	0.075	0.467	0
zpm25	0.526	0.277	0.112	-0.564	-0.563	0
zso2	0.424	-0.328	0.797	0.206	0.186	0

Table 4: Regression Results

	agric	food	soda	beer	smoke	toys
mrf	0.771^{***}	0.823^{***}	0.686^{***}	0.724^{***}	0.733^{***}	0.990^{***}
smb	0.0684	-0.177^{***}	-0.353^{***}	-0.368^{***}	-0.329^{***}	0.433^{***}
hml	0.0439	-0.266^{***}	-0.284^{***}	-0.303^{***}	-0.319^{***}	-0.217^{**}
rmw	0.137	0.416^{***}	0.423^{***}	0.422^{***}	0.558^{***}	0.331^{***}
cma	0.178	0.803^{***}	0.738^{***}	0.633^{***}	0.804^{***}	0.252^{*}
pc1	0.0280	0.0302^{**}	-0.00342	0.00232	0.00774	-0.00390
_cons	-0.0155	-0.00306	0.00659	0.0187	-0.00119	-0.00947
N	1617	1617	1617	1617	1617	1617
R^2	0.308	0.506	0.373	0.476	0.296	0.375

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Regression Results Continued

	fun	books	hshld	hlth	medeq	drugs
mrf	1.193***	1.027***	0.809***	0.849***	0.895***	0.936***
smb	0.156**	0.637***	-0.230***	0.373***	0.0342	-0.0859**
hml	-0.339***	0.0959*	-0.276***	-0.263***	-0.589***	-0.641***
rmw	-0.284***	0.284***	0.409***	0.0336	-0.182***	-0.570***
cma	-0.348***	0.225**	0.822***	0.139	0.187***	0.314***
pc1	-0.0218	-0.0101	0.00382	0.00608	0.00123	0.00391
_cons	0.0211	-0.0218	0.00214	0.00186	0.0250	0.00175
N	1617	1617	1617	1617	1617	1617
R^2	0.584	0.659	0.587	0.448	0.713	0.716

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Regression Results Continued

	chems	rubbr	txtls	bldmt	cnstr	steel
mrf	1.106***	0.965***	1.094***	1.170***	1.180***	1.379***
smb	0.145***	0.389***	0.363***	0.587***	0.647***	0.758***
hml	0.172***	-0.0959**	-0.103	0.118***	0.232***	0.552***
rmw	0.122**	0.236***	0.391***	0.292***	0.246***	0.0429
cma	0.287***	0.278***	-0.00806	0.354***	0.218**	0.581***
pc1	0.00651	0.00375	0.0222	-0.00296	0.0114	-0.00451
_cons	-0.0110	0.00683	-0.00813	-0.00777	-0.0122	-0.0327
N	1617	1617	1617	1617	1617	1617
R^2	0.725	0.689	0.439	0.801	0.636	0.622

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Regression Results Continued

	fabpr	mach	elceq	autos	aero	ships
mrf	1.234***	1.222***	1.162***	1.159***	1.077***	1.198***
smb	1.088***	0.339***	0.395***	0.415***	-0.0168	0.686***
hml	0.472***	0.252***	0.133***	0.271***	-0.0112	0.232***
rmw	0.121	0.206***	0.151***	0.229***	0.308***	0.197**
cma	0.181	0.251***	0.423***	-0.0223	0.302***	0.435***
pc1	0.0145	-0.00198	0.00320	-0.00950	0.00287	0.0212
_cons	-0.0122	-0.00997	-0.0180	-0.0216	0.0208	0.0136
N	1617	1617	1617	1617	1617	1617
R^2	0.488	0.798	0.781	0.667	0.637	0.580

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Regression Results Continued

	guns	gold	mines	coal	oil	util
mrf	0.836***	0.448***	1.332***	1.368***	1.151***	0.616***
smb	-0.0451	0.193	0.401***	0.982***	-0.0359	-0.243***
hml	-0.262***	-0.183	0.519***	1.054***	0.514***	-0.220***
rmw	0.323***	-0.314	-0.0821	-0.629**	-0.755***	0.214***

cma	0.451^{***}	1.207^{***}	0.551^{***}	0.852^{***}	0.754^{***}	0.733^{***}
pc1	-0.00557	-0.0239	-0.0120	-0.0133	-0.00600	0.00955
_cons	0.0531^{**}	-0.00829	-0.0465	-0.0926	-0.0358	0.0134
N	1617	1617	1617	1617	1617	1617
R^2	0.395	0.038	0.497	0.272	0.629	0.324

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Regression Results Continued

	telcm	persv	bussv	hardw	softw	chips
mrf	0.878^{***}	1.039^{***}	1.050^{***}	1.104^{***}	1.047^{***}	1.093^{***}
smb	-0.0551^{*}	0.577^{***}	0.186^{***}	0.0597	-0.126^{***}	0.00641
hml	-0.0347	0.0882^{*}	-0.170^{***}	-0.0661	-0.288^{***}	-0.0998^{*}
rmw	0.272^{***}	0.240^{***}	0.0752^{**}	0.0737	-0.137^{***}	0.310^{***}
cma	0.441^{***}	0.191^{**}	-0.0158	-0.189^{*}	-0.729^{***}	-0.710^{***}
pc1	0.00214	-0.0208^{*}	-0.00124	-0.0147	-0.00748	-0.000746
_cons	0.00196	-0.00179	0.0201^{**}	-0.0103	0.00918	0.0161
N	1617	1617	1617	1617	1617	1617
R^2	0.639	0.708	0.899	0.651	0.858	0.713

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Regression Results Continued

	labeq	paper	boxes	trans	whlsl	rtail
mrf	1.121^{***}	1.029^{***}	1.037^{***}	1.133^{***}	0.964^{***}	0.982^{***}
smb	0.0932^{***}	-0.00293	0.203^{***}	0.233^{***}	0.326^{***}	0.00937
hml	-0.297^{***}	-0.0572	0.0636	0.117^{**}	-0.0271	-0.289^{***}
rmw	-0.0845^{*}	0.419^{***}	0.279^{***}	0.436^{***}	0.192^{***}	0.452^{***}
cma	0.171^{***}	0.504^{***}	0.425^{***}	0.196^{**}	0.327^{***}	0.0247
pc1	-0.00295	-0.000334	-0.00471	0.0154	0.0130^{*}	0.00675
_cons	0.0177	-0.00635	0.00183	0.00393	-0.00459	0.00710
N	1617	1617	1617	1617	1617	1617
R^2	0.835	0.744	0.587	0.718	0.811	0.765

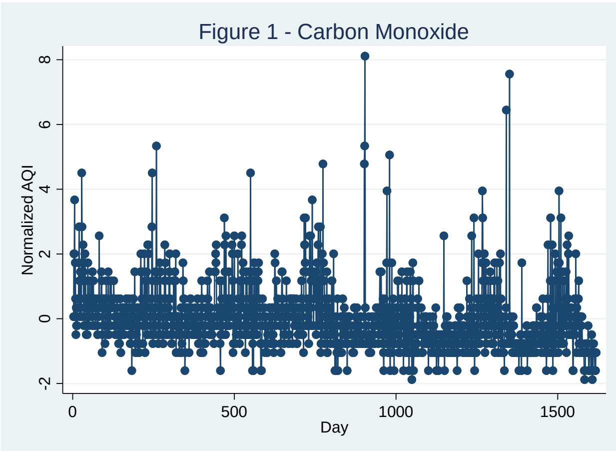
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

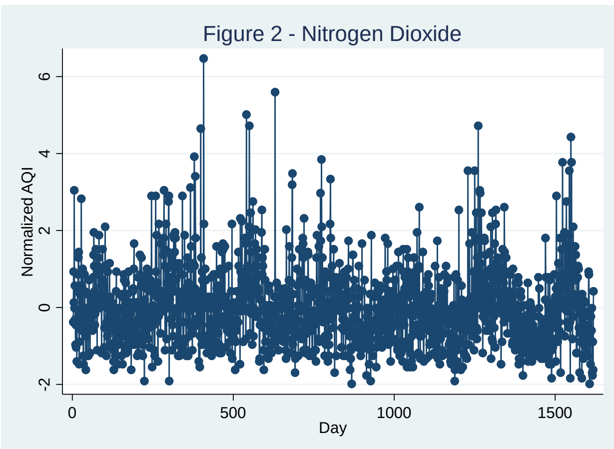
Table 4: Regression Results Continued

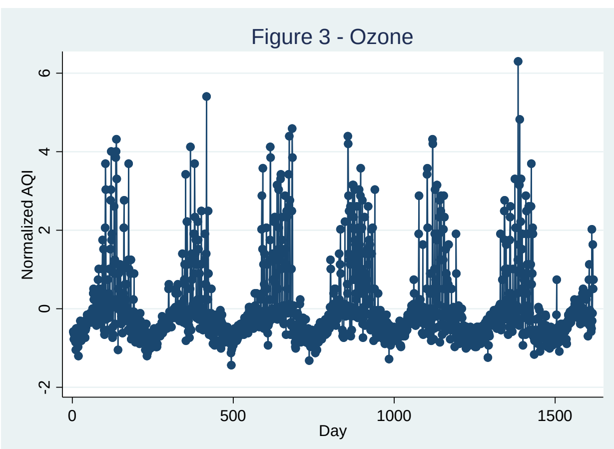
	meals	banks	insur	rlest	fin	other
mrf	0.821^{***}	1.059^{***}	0.922^{***}	1.098^{***}	1.178^{***}	0.959^{***}
smb	-0.0271	0.0913^{***}	-0.0377	0.446^{***}	0.0909^{***}	-0.203^{***}
hml	-0.159^{***}	1.025^{***}	0.339^{***}	0.125^{**}	0.765^{***}	0.274^{***}
rmw	0.259^{***}	-0.327^{***}	-0.0614	0.193^{***}	-0.329^{***}	-0.108^{**}
cma	0.124^{*}	-0.718^{***}	-0.103^{*}	0.0348	-0.403^{***}	0.255^{***}
pc1	-0.00404	-0.0117	-0.00114	0.0108	-0.0139	0.00631
_cons	0.0210	0.0142	0.0274^{**}	-0.0270	0.00898	-0.00795
N	1617	1617	1617	1617	1617	1617
R^2	0.625	0.879	0.780	0.674	0.859	0.779

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix 2: Figures







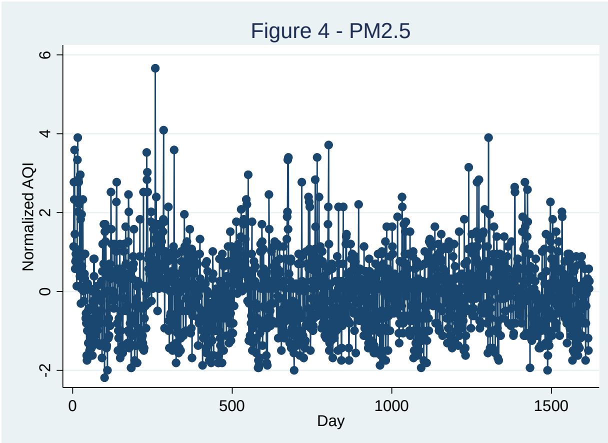


Figure 4 - PM2.5
Normalized AQI
-2
0
2
4
6
0
500
1000
1500
Day

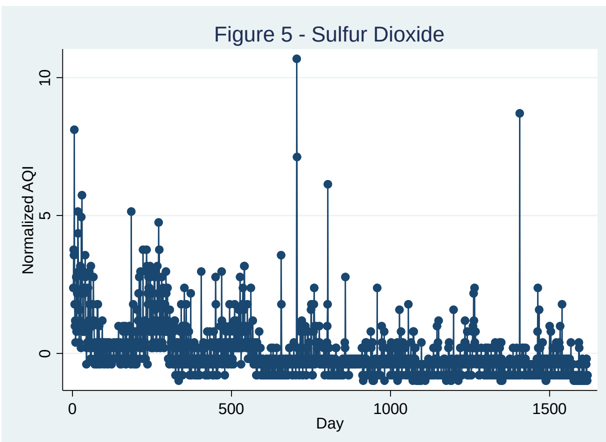


Figure 5 - Sulfur Dioxide
Normalized AQI
0
5
10
0
500
1000
1500
Day

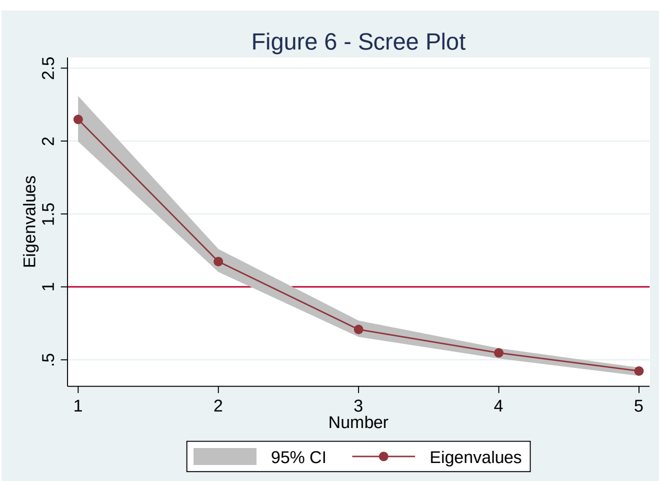


Figure 6 - Scree Plot
Eigenvalues
.5
1
1.5
2
2.5
1
2
3
4
5
Number
95% CI
Eigenvalues

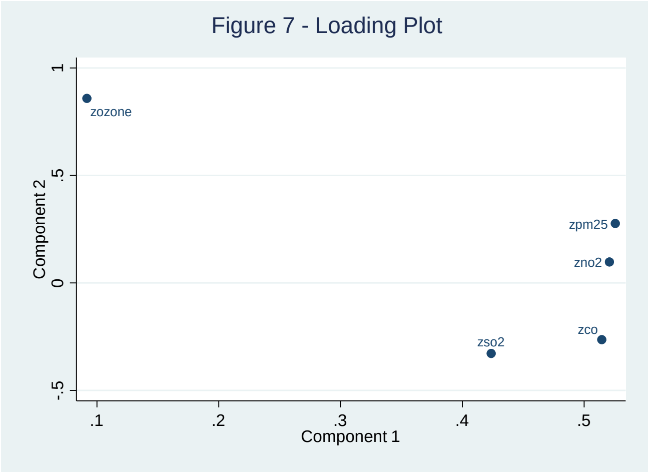


Figure 7 - Loading Plot
Component 2
1
.5
0
-.5
.1
.2
.3
.4
.5
Component 1
zozone
zpm25
zno2
zso2
zco

Chapter 4: Sample Selection Bias in Hedonic Pricing Models of Thoroughbred Broodmares

Abstract

An issue with OLS estimation in hedonic pricing model literature is that they do not account for sample selection bias. In broodmare auctions, the purchase decision and whether a price is realized or zero is endogenous. This paper contributes to the hedonic broodmare price analysis literature by implementing the Heckman (1976) sample selection regression to estimate a hedonic pricing model using data from the 2020 January Keenland Sale. Many published papers do not accommodate this selection process and may have biased coefficients. This paper further contributes methodologically to the thoroughbred broodmare literature by estimating a Bayesian Heckman model. The sire's stud fee, domestic status, and the day of the session are significant for broodmare prices. This may be implemented within a profit maximizing purchasing and breeding strategy.

Keywords

Hedonic models, sample selection bias, thoroughbred broodmares, Bayesian methods

Introduction

American Pharoah (2015) and Justify (2018) recently claimed the Triple Crown of Thoroughbred Racing after a draught since Affirmed (1978) took the title. Thirteen racers have won the prestigious award in history, with some earning more than $10 million in today's dollars.

The American Horse Council Foundation estimates 7.2 million horses are in the United States consuming 32 million acres of owned land and another 49 million acres of leased land. Further, they estimate that the direct effect of the horse industry on the domestic economy is approximately $50 billion. The direct employment total reaches near one million jobs earning roughly $38 million in various accounts. The ripple effect from this gigantic industry is estimated to be $122 billion impact and 1.7 million jobs, respectively. The high stakes associated with thoroughbred horseracing makes understanding the determinants of prices economically important for both buyers and sellers. According to Chizum and Wimmer (1997) and Wimmer and Chizum (2006), asymmetric information and adverse selection prevail in Thoroughbred markets. These issues may make the empirical findings of this paper useful for increasing market efficiency.

Vickner (2018) points out that among all the hedonic price models applied to Thoroughbreds, a majority study yearling. Only Neibergs (2001), Maynard and Stoeppel (2007), and Dority et al. (2016) focus on broodmares. Chezum and Wimmer (1997), Vickner and Koch (2001), Robbins and Kennedy (2001) Wimmer and Chezum (2006), Parson and Smith (2008), Plant and Stowe (2013), Marion and Stowe (2016) all focus predominately on yearlings. Stowe and Ajello (2010) perform OLS in their hedonic pricing model of stud fee determinants, while Stowe (2013) extends this model to include fixed effects. Taylor et al. (2006) uses the Heckman model within the horse literature on data about quarter horses. This paper contributes to the relatively scarce literature on broodmare pricing by applying the Heckman model to account for sample selection bias. Failure to consider the endogenous selection process causes estimates to be bias and would misinform prospective buyers, sellers, bloodline agents, policymakers, and fellow scientists.

Vickner (2018) suggests that a Bayesian Heckman selection model would be an interesting future contribution to the thoroughbred literature. Heckman et al. (2013) provide the details for extending the classical selection model to a Bayesian framework. Ng'ombe and Boyer (2019) point out that a Bayesian inference is desirable as it is exact for any sample size.

Methodology

A Heckman sample selection framework is applied to a hedonic pricing model for broodmares at the 2020 January Keenland Sale.

$$1)\quad \mathrm{y} = \sum_{\mathrm{b}=1}^{5} \beta_{\mathrm{b}} \mathrm{x}_{\mathrm{b}} + \mathrm{x}_{\mathrm{g}} + \sum_{\mathrm{m}=1}^{4} \beta_{\mathrm{m}} \mathrm{x}_{\mathrm{m}} + \mathrm{u}_1$$

where y is the natural log of broodmare selling price, x_b represents breeding characteristics, and x_g, and x_m are genetic, and market characteristics, respectively. The error term u_1 is

$$2)\quad \mathrm{u}_1 \sim \mathrm{N}(0, \sigma)$$

The covariates in this model are motivated by Maynard and Stoeppel (2007) and Dority et al. (2016). They argue that breeding, genetic, and market characteristics are relevant in explaining thoroughbred broodmare auction prices. The breeding characteristics in the model are a dummy = 1 if a broodmare is a prospect, the age in years, color dummy=1 if the broodmare is black, sire earning, and sire stud fee. Stowe (2013) finds that sire stud fee is highly explained by progeny sale price.

Poerwanto and Stowe (2010) find a positive relationship between the number of foals produced by a sire and sire's yearlings' average selling price, therefore sire representation is included into the model as a genetic characteristic. Market characteristics include a dummy = 1 if the sire is domestic and dummies for the days of the auction. In the case of broodmare auctions, each individual broodmare is not sold, and the price is only observed if the selection equation is satisfied:

$$3)\quad \gamma_{\mathrm{s}} \mathrm{z}_{\mathrm{s}} + \mathrm{u}_2 > 0$$

where z_s is a dummy =1 if the sire has won a Triple Crown race. This can be the Kentucky Derby, Preakness Stakes, or Belmont Stakes. The error term u_2 is

4) $u_2 \sim N(0,1)$

5) $corr(u_{1,} u_2) = \rho$

Estimating hedonic pricing models via OLS in the existence of this error correlation causes estimates to be biased, as it violates the assumption of random sampling. Dority et al. (2016) does not account for sample selection processes. Heteroskedasticity may also arise. Maynard and Stoeppel (2007) account for heteroskedasticity using a Box-Cox transformation. Marion and Stowe (2016) use a Breusch Pagan test and reject the null hypothesis of heteroskedasticity.

The methodological contribution of this paper is the Bayesian Heckman model applied to thoroughbred broodmare auctions. The Random-walk Metropolis-Hastings algorithm was used. Markov Chain Monte Carlo (MCMC) sample size is 50,000. 20,000 burn-ins were used.

Data and Descriptive Statistics

Data on broodmare sales price and characteristics were obtained from the 2020 January Keenland Sale at Keenland Association in Lexington, KY. The sire nationality and performance data were obtained from the Blood-Horse Stallion Register and matched to corresponding broodmares. Table 1 presents the descriptive statistics.

The sample contains 524 unique broodmares. 323 (61.6%) of those ended up being sold. The other sale prices are recorded as zero. The average price conditional on being sold is $44,889 and ranges from $1,000 to $640,000. Broodmare prospects account for roughly 51% of the sample and average prospects have a price of $28,079 versus $27,241 of the average non-prospects. The difference, however, is statistically insignificant, with a p-value of 0.879 as shown in Figure 1.

The average broodmare in the sample is approximately 6 years of age. The average sire earned $2.08 million, has a stud fee of about $63.81, and is being represented 10 times. 95% of sires are domestic, and 16.8% total sire have won a Triple Crown race. A broodmare of a domestic sire on average sold for $28,379.92 versus 14,076.92. The difference is not statistically significant, with a p-value of 0.261. Figure 2 show this comparison.

Table 2 presents a pairwise comparison of mean price across different auction sessions. There are statistically significant differences in price between session 1 to sessions 3, 4, and 5, respectively as well as between 2 and 3, 4, and 5, respectively. The signs on each of these differences are negative and have management implications. Buyers may be able to receive a discounted price if they are willing to delay their purchase by attending a later auction. This inference is consistent whether using Bonferroni, Sidek, Sheffe, Tukey, SNK, Duncan, or Dunnet adjustment. Figure 3 visualizes this relationship. Dority et al. (2016) find that the longer buyers are willing to wait, the lower price that they can receive.

Results and Discussion

Table 3 presents the posterior summary statistics from the Bayesian Heckman model performed in Stata. The sire's stud fee, domestic status, and the session are found to have a statistically significant effect on average broodmare selling price.

A one percent increase in sire's stud fee is expected to increase the selling price of the broodmare .23 percent on average holding other variables constant. This value is like Neiberg (2001) who find 0.21 and less than the 0.65 of Dority et al. (2016). This provides further evidence for the necessity to model sample selection processes. The corrected model gives managers better estimates of possible returns to sire earnings in the breeding market. Figure 4 shows the diagnostic plots for the stud fee variable. Autocorrelation is shown to decline in the initial lags, indicating convergence of the model. A domestic sire is associated with an 86.6% on average broodmare selling price relative to non-domestic all else equal. For managers, this means that including a domestic sire in your bloodline may increase future broodmare returns. Figure 5 shows the diagnostic plots for the domestic status variable. Autocorrelation is shown to decline in the initial lags.

The result on sire representation also has management relevance. When deciding the number of mares for a sire to service, managers must trade-off short term earnings with the possibility of decreasing future value of the sire due to the possibility of inadequate foal. Based on the statistically insignificant results, the relationship between the sire representation and broodmare price is inconclusive.

The dummies indicating the session are all statistically significant except for the session 2 dummy. This is consistent with Dority et al. (2016) who find what they describe as buyer fatigue in these auctions. They point out that it is customary for the highest quality broodmares to be auctioned the earliest and that there may be a psychological notion that the best lot has been sold. This evidence suggests potential buyers who wait until later auction sessions incur additional risk. Managers and potential buyers should seek to attend the earliest sessions.

Conclusion

This paper contributes to the thoroughbred literature by estimating a Heckman sample selection model to 2020 January Keenland Sales data to account for the sample selection process underlying broodmare sales. Given the documented asymmetric information and adverse selection in the Thoroughbred industry, an unbiased hedonic pricing model of broodmares stands to inform buyers of the characteristics important in determining price. This evidence may alleviate some inefficiency associated with the information gap and market failure. In an industry with roughly $175 billion economic impact, the welfare loss from this inefficiency is likely nontrivial. Failure to account for the selection process prevalent by omitting broodmares with prices of zero from the sample will bias coefficient estimates and misinform prospective buyers, breeders, and racers. This estimation procedure, combined with the exactness of Bayesian inference, can be used in future Thoroughbred hedonic pricing analyses, whether for broodmares or yearlings.

Sire's stud fee, domestic status, and the day of the auction session are all statistically significant factors in broodmare prices. Managers can implement this information into their buying and breeding strategies. Further studies may examine other variables, such as dam characteristics, sprinting speed, or breeder characteristics for significance, but should be aware of the modelling issues addressed in this paper.

References

Chezum, Brian, and Brad Wimmer. "Roses or lemons: adverse selection in the market for thoroughbred yearlings." *Review of Economics and Statistics* 79, no. 3 (1997): 521-526.

Dority, Bree L., Dayna Larreau, and Frank Tenkorang. "Hedonic Price Analysis of Non-Barren Broodmares." *Review of Economic and Business Studies* 9, no. 2 (2016): 61-75.

Heckman, James J. "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models." In *Annals of economic and social measurement, volume 5, number 4*, pp. 475-492. NBER, 1976.

Heckman, James J., Hedibert F. Lopes, and Rémi Piatek. "Treatment effects: A Bayesian perspective." *Econometric reviews* 33, no. 1-4 (2014): 36-67.

Robert, Marion, and C. Jill Stowe. "Ready to run: price determinants of thoroughbreds from 2 year olds in training sales." *Applied Economics* 48, no. 48 (2016): 4690-4697.

Maynard, Leigh J., and Kelly M. Stoeppel. "Hedonic price analysis of thoroughbred broodmares in foal." *Journal of Agribusiness* 25, no. 345-2016-15149 (2007): 181-195.

Neibergs, J. Shannon. "A hedonic price analysis of thoroughbred broodmare characteristics." *Agribusiness: An International Journal* 17, no. 2 (2001): 299-314.

Ng'ombe, John N., and Tracy A. Boyer. "Determinants of earthquake damage liability assignment in Oklahoma: A Bayesian Tobit censored approach." Energy Policy 131 (2019): 422-433.

Parsons, Conor, and Ian Smith. "The price of thoroughbred yearlings in Britain." *Journal of Sports Economics* 9, no. 1 (2008): 43-66.

Plant, Emily J., and C. Jill Stowe. "The price of disclosure in the thoroughbred yearling market." *Journal of Agricultural and Applied Economics* 45, no. 2 (2013): 243-257.

Poerwanto, Devie A., and C. Jill Stowe. "The relationship between sire representation and average yearling prices in the Thoroughbred industry." *Journal of Agribusiness* 28, no. 345-2016-15276 (2010): 61-74.

Robbins, Mark, and Peter E. Kennedy. "Buyer behaviour in a regional thoroughbred yearling market." *Applied Economics* 33, no. 8 (2001): 969-977.

Stowe, C. Jill. "Breeding to sell: a hedonic price analysis of leading Thoroughbred sire stud fees." *Applied Economics* 45, no. 7 (2013): 877-885.

Stowe, C. Jill, and Billy Ajello. "A Hedonic price analysis of differentiated products of unknown quality: Freshman sire stud fees in the thoroughbred breeding industry." *Journal of Agribusiness* 28, no. 345-2017-1098 (2010): 19-30.

Taylor, Mykel R., Kevin C. Dhuyvetter, Terry L. Kastens, Megan Douthit, and Thomas L. Marsh. "Show quality quarter horse auctions: price determinants and buy-back practices." *Journal of Agricultural and Resource Economics* (2006): 595-615.

Vickner, Steven S., and Stephen I. Koch. "Hedonic pricing, information, and the market for thoroughbred yearlings." *Journal of Agribusiness* 19, no. 345-2016-15190 (2001): 173-189.

Wimmer, Bradley S., and Brian Chezum. "Adverse selection, seller effort, and selection bias." *Southern Economic Journal* (2006): 201-218.

Appendix 1: Tables

Table 1: Descriptive Statistics

Variable	Mean	Std.Dev.	Min	Max
price	27670.23	63181.87	0	640000
prospect	.511	.5	0	1
age in years	5.742	2.519	2	16
black	.468	.499	0	1
sire stud fee	63812.34	67347.84	2000	250000
sire earnings	2080000	2260000	32400	1.05e+07
representation	10.439	6.488	1	25
domestic	.95	.217	0	1
Triple Crown	.168	.374	0	1
# Observations	524			

Table 2: Pairwise comparisons of mean price across sessions with Bonferroni adjustment

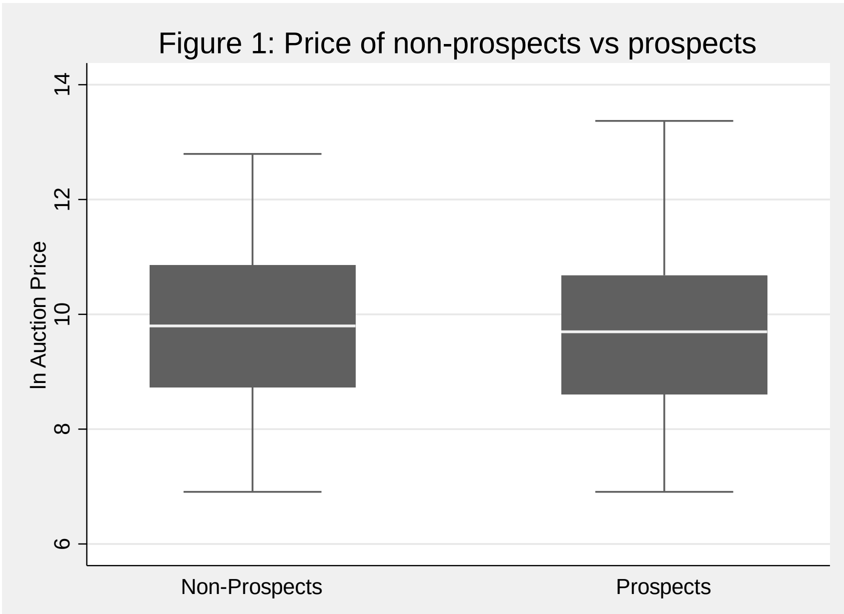
Session	Contrast	Std.Err.	P>t
2_vs_1	6935.513	7710.558	1.000
3_vs_1	-21575.94	8390.539	0.104
4_vs_1	-28729.55	8591.845	0.009
5_vs_1	-31674.73	8858.497	0.004
3_vs_2	-28511.46	8122.274	0.005
4_vs_2	-35665.06	8330.065	0.000
5_vs_2	-38610.25	8604.833	0.000
4_vs_3	-7153.605	8963.17	1.000
5_vs_3	-10098.79	9219.088	1.000
5_vs_4	-2945.185	9402.672	1.000

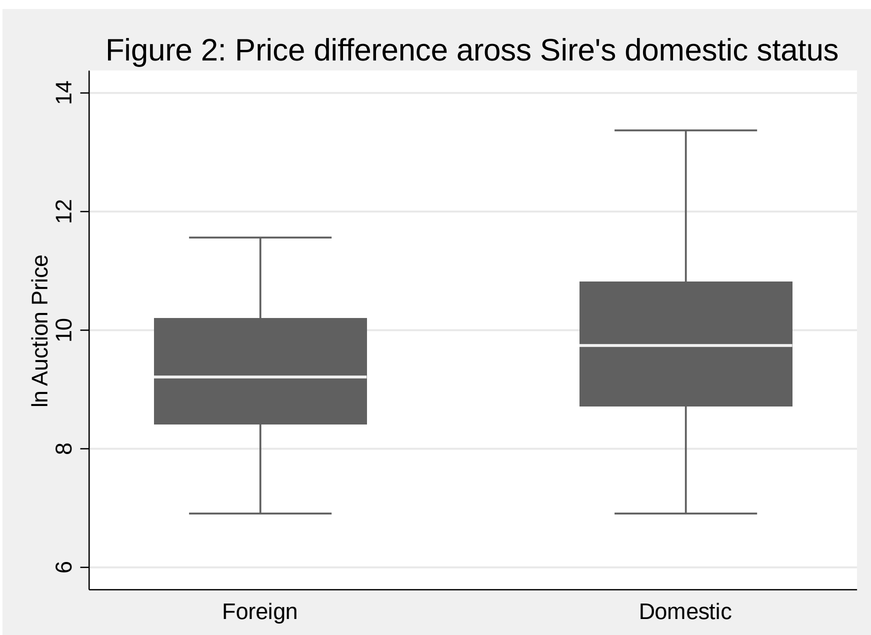
Table 3: Bayesian Heckman Posterior Summary Statistics

MCMC sample size = 50,000, Burn-in = 20,000

	Mean	Std.Dev.	MCSE	Median	[95%	Cred.]
lnp						
prospect	0.097	0.106	0.007	0.094	-0.105	0.318
ageyears	-0.046	0.030	0.001	-0.046	-0.105	0.012
black	0.190	0.140	0.011	0.189	-0.084	0.465
lnsireearnings	0.028	0.054	0.004	0.028	-0.077	0.132
lnsirefee	0.233	0.078	0.005	0.231	0.083	0.387
representation	0.006	0.015	0.001	0.007	-0.026	0.035
siredomestic	0.866	0.223	0.012	0.862	0.432	1.303
session						
2	0.145	0.105	0.007	0.146	-0.063	0.342
3	-0.813	0.169	0.018	-0.814	-1.143	-0.479
4	-1.412	0.161	0.010	-1.416	-1.724	-1.096
5	-1.569	0.192	0.011	-1.568	-1.949	-1.194
_cons	7.395	1.046	0.077	7.406	5.252	9.382
select						
siretcwinner	-0.086	0.140	0.005	-0.091	-0.348	0.201
_cons	0.312	0.060	0.001	0.312	0.193	0.429
athrho	-0.910	0.473	0.058	-1.032	-1.500	0.446
lnsigma	0.389	0.097	0.010	0.399	0.187	0.556

Appendix 2: Figures





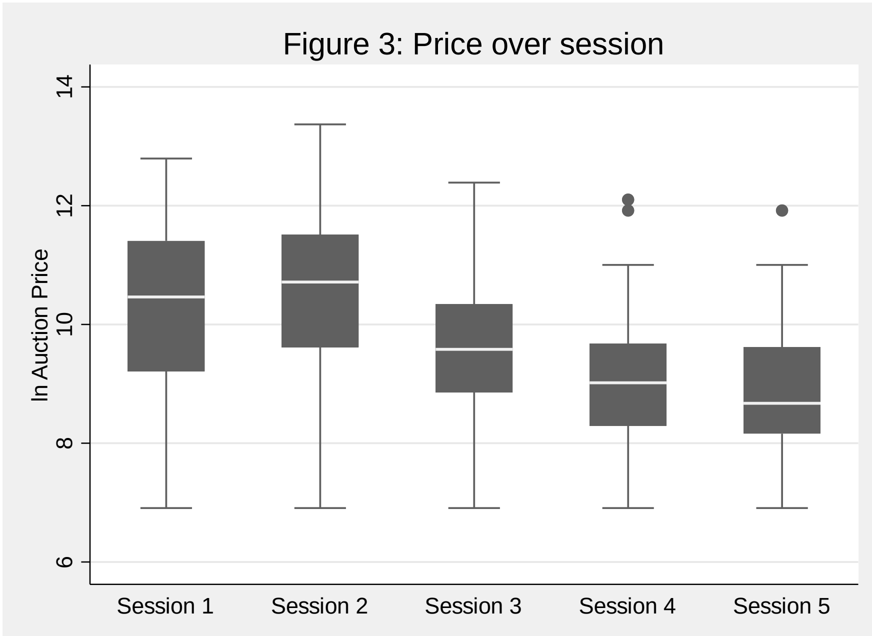


Figure 3: Price over session

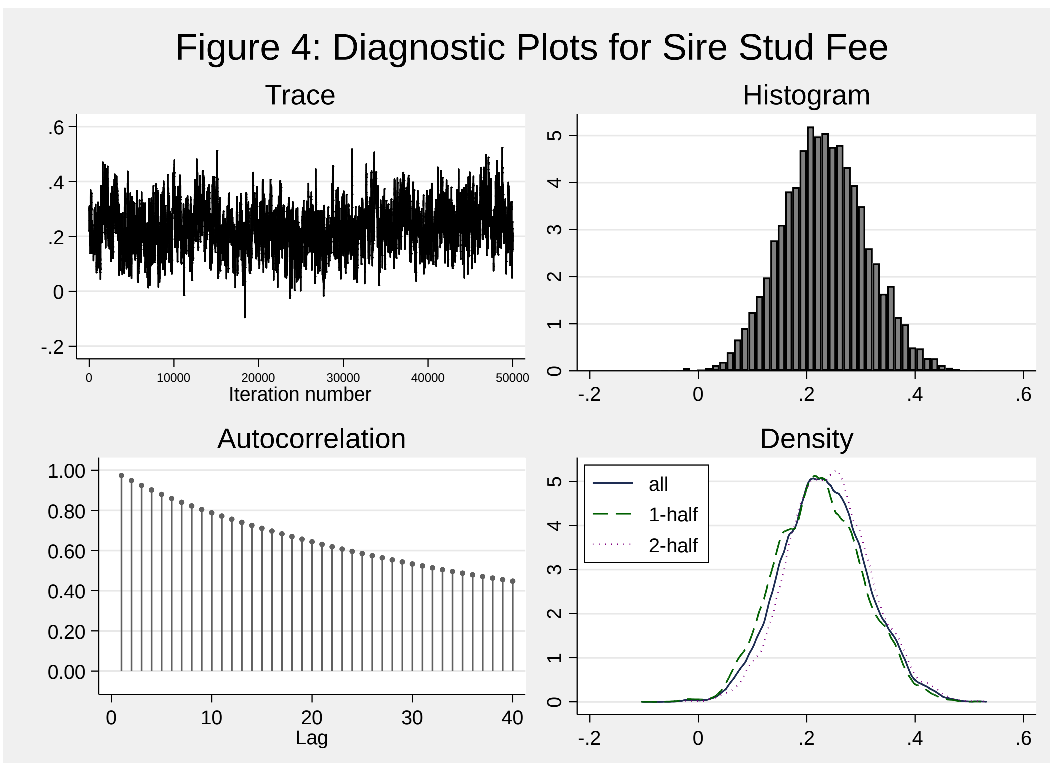


Figure 4: Diagnostic Plots for Sire Stud Fee

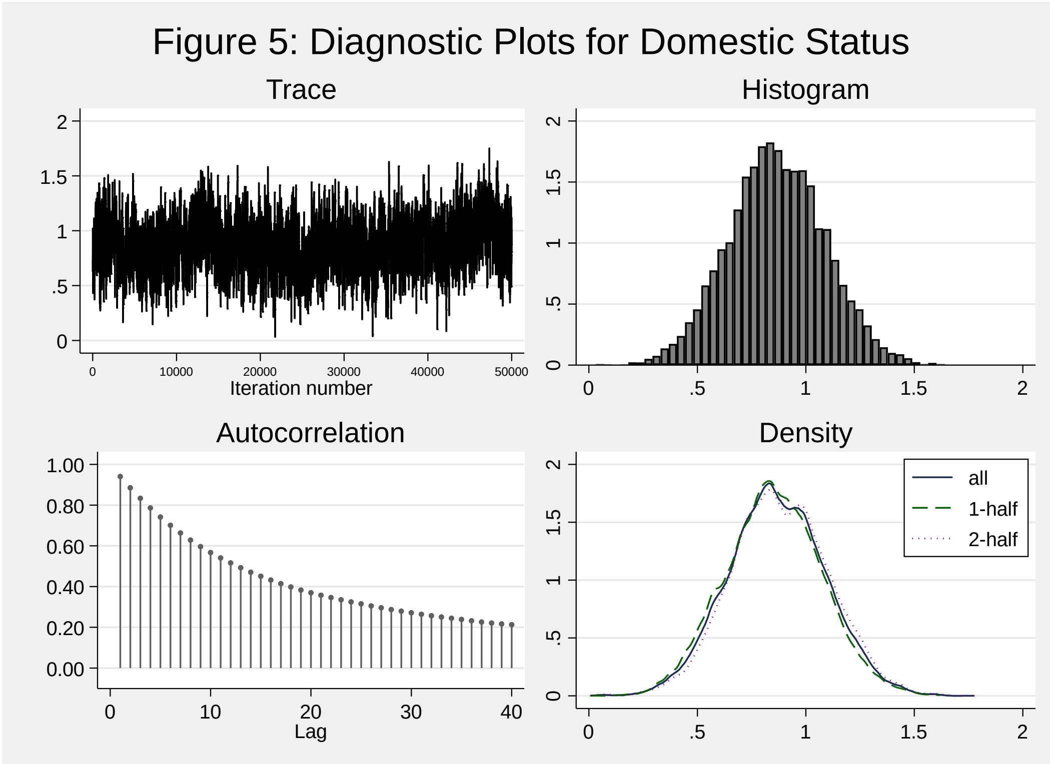


Figure 5: Diagnostic Plots for Domestic Status
Trace
Histogram
Autocorrelation
Density
Iteration number
Lag
all
1-half
2-half