LI, SHU. Three Essays on Agricultural Commodity Markets and Barge Transportation on the Mississippi Waterways. (Under the direction of Walter Thurman.)

The first essay analyzes the effects of waterway transportation costs on the spatial distribution of corn prices at U.S. grain markets. This paper develops theory-based predictions on how effects of transportation costs on corn prices can depend on the geographical locations and characteristics of markets. The spatial effects of barge rates on prices are measured and tested by applying a mixture of parametric and nonparametric methods to a rich panel data set of corn prices from over one thousand locations. The second essay studies the information content of China’s soybean futures market. This paper tests conditional and unconditional unbiasedness of futures price as a forecast of future spot price. It compares and measures the out-of-sample fitting errors of futures prices with other econometric models in terms of Average Percentage Error (APE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Percentage Error (RMSPE). The third essay is a study of the relationship between water flow on the Mississippi river and grain barge transportation costs. The paper studies the effect of a unit change of water level in the Lower Mississippi river on barges rates for transporting grain from St. Louis to export ports in the New Orleans Region.
© Copyright 2013 by Shu Li

All Rights Reserved
Three Essays on Agricultural Commodity Markets and Barge Transportation on the Mississippi Waterways

by
Shu Li

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

in Economics

Raleigh, North Carolina

2013

APPROVED BY:

__________________________    ____________________________
Kelly Zering             Denis Pelletier

__________________________    ____________________________
Yichao Wu                 Walter Thurman
Chair of Advisory Committee
Dedication

To my parents.
Biography

The author was born in Inner-Mongolia in China. She has earned her bachelor’s degree in insurance from Central University of Finance and Economics in 2005. She has earned her master’s degree in statistics from Stephen F. Austin State University in 2008. She began her Ph.D. degree in economics at North Carolina State University since 2008.
Acknowledgements

I would like to thank my advisor Dr. Walter Thurman for his advice. I also thank my committee members, Dr. Kelly Zering, Dr. Denis Pelletier and Dr. Yichao Wu for their helpful comments and advice.
# Table of Contents

List of Tables ................................................................. vii

List of Figures ................................................................. viii

Chapter 1 Grain Transport on the Mississippi River and Spatial Corn Basis ................................................. 1
  1.1 Introduction ............................................................. 1
  1.1.1 Corn Production and Exports ................................... 1
  1.1.2 Corn Transportation ............................................. 3
  1.2 Data .................................................................. 10
  1.3 An Econometric Model of Corn Basis and Transportation Cost .................................................. 13
  1.4 Estimates ............................................................... 19
    1.4.1 Regression Results .......................................... 19
    1.4.2 Smoothing and Mapping .................................... 25
  1.5 Concluding Remarks ................................................. 35

Chapter 2 The Information Content of Futures Prices in China’s Soybean Futures Market .............................. 37
  2.1 Introduction ............................................................. 37
  2.2 Literature Review .................................................... 39
  2.3 Methodology ........................................................... 42
  2.4 Data .................................................................. 46
  2.5 Estimation ............................................................... 47
    2.5.1 Out-of-Sample Fits .......................................... 56
    2.5.2 Conclusion ..................................................... 57

Chapter 3 Water Levels on the Mississippi River and Grain Transport Barge Rates ........................................... 58
  3.1 Introduction ............................................................. 58
  3.2 Literature Review .................................................... 60
  3.3 Models ................................................................. 61
  3.4 Data .................................................................. 64
List of Tables

Table 1.1  Forward Selection: Proportion of regressions that are significant at 10% level ................................. 14
Table 1.2  Forward Selection: Proportion of regressions that are significant at 5% level ........................................ 15
Table 1.3  Backward Selection: Proportion of regressions that are significant at 10% level ................................. 17
Table 1.4  Backward Selection: Proportion of regressions that are significant at 5% level ........................................ 18
Table 1.5  Statistics and tests on barge rates coefficients .......................................................... 22
Table 2.1  Descriptive statistics of futures prices of expiring contracts ................................. 48
Table 2.2  Estimated OLS Regression for Model (2.2) .......................................................... 48
Table 2.3  Estimation of Model (2.7) .................................................................................. 49
Table 2.4  Dickey-Pantula tests ......................................................................................... 50
Table 2.5  The augmented Dickey-Fuller Unit Root Tests for 2-month-ahead Futures Prices .......................................................... 51
Table 2.6  The Phillips-Perron Unit Root Test for 2-month-ahead Futures Price ................................. 52
Table 2.7  Unconditional Unbiasedness Tests under Cointegration .......................................................... 53
Table 2.8  Estimates of Random Walk Model 2.8 ........................................................................ 53
Table 2.9  Estimated Error Correction Model ........................................................................... 55
Table 2.10 Out-of-Sample Fit Comparison .......................................................... 56
Table 3.1  Descriptive Statistics of Barge Rate at St. Louis .......................................................... 65
Table 3.2  Descriptive Statistics of Water Levels at MISL and OHCA .......................................................... 66
Table 3.3  Estimates of Model 3.3 and 3.4 .................................................................................. 74
Table 3.4  Model with only contemporaneous water levels .......................................................... 78
Table 3.5  Model with up to 1 lagged water levels ........................................................................... 79
Table 3.6  Model with up to 2 lagged water levels ........................................................................... 80
Table 3.7  “Long-run” Effects of Water Levels at MISL and OHCA .......................................................... 81
### List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Corn surplus/deficit map with the transportation system</td>
<td>2</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>Corn modal shares</td>
<td>3</td>
</tr>
<tr>
<td>Figure 1.3</td>
<td>2007 barge tonnage, by major commodity group</td>
<td>4</td>
</tr>
<tr>
<td>Figure 1.4</td>
<td>Agriculturally-significant waterways</td>
<td>5</td>
</tr>
<tr>
<td>Figure 1.5</td>
<td>Agricultural and total freight moving on U.S. waterways</td>
<td>5</td>
</tr>
<tr>
<td>Figure 1.6</td>
<td>Prices and Grain Flow from Market i</td>
<td>7</td>
</tr>
<tr>
<td>Figure 1.7</td>
<td>The Rivers</td>
<td>11</td>
</tr>
<tr>
<td>Figure 1.8</td>
<td>All Markets</td>
<td>11</td>
</tr>
<tr>
<td>Figure 1.9</td>
<td>Corn Markets within 150 miles of major navigable rivers</td>
<td>12</td>
</tr>
<tr>
<td>Figure 1.10</td>
<td>Estimated Seasonal Factor for Markets #2551, #1423, #1014, #4796 and #4092</td>
<td>20</td>
</tr>
<tr>
<td>Figure 1.11</td>
<td>Markets #4796, #2551, #4092, #1423 and #1014 from North to South</td>
<td>20</td>
</tr>
<tr>
<td>Figure 1.12</td>
<td>Histogram of $\beta_0$</td>
<td>23</td>
</tr>
<tr>
<td>Figure 1.13</td>
<td>Histogram of $\beta_1$</td>
<td>24</td>
</tr>
<tr>
<td>Figure 1.14</td>
<td>Histogram of $\beta_2$</td>
<td>24</td>
</tr>
<tr>
<td>Figure 1.15</td>
<td>Histogram of long-run effect: sum of $\beta_i$’s</td>
<td>25</td>
</tr>
<tr>
<td>Figure 1.16</td>
<td>Histogram of $\gamma_0$</td>
<td>26</td>
</tr>
<tr>
<td>Figure 1.17</td>
<td>Contour Map of Barge Rate Effects Using Kernel Smoothing (SD = 50 miles)</td>
<td>28</td>
</tr>
<tr>
<td>Figure 1.18</td>
<td>Contour Map of Barge Rate Effects Using Kernel Smoothing (SD = 50 miles)</td>
<td>28</td>
</tr>
<tr>
<td>Figure 1.19</td>
<td>Contour Map of Barge Rate Effects Using Kernel Smoothing (SD = 30 miles)</td>
<td>29</td>
</tr>
<tr>
<td>Figure 1.20</td>
<td>Contour Map of Barge Rate Effects Using Kernel Smoothing (SD = 70 miles)</td>
<td>29</td>
</tr>
<tr>
<td>Figure 1.21</td>
<td>Contour Map of Barge Rate Effects Using loess Smoother</td>
<td>30</td>
</tr>
<tr>
<td>Figure 1.22</td>
<td>Contour Map of Barge Rate Effects Using loess Smoother w/ Markets</td>
<td>31</td>
</tr>
<tr>
<td>Figure 1.23</td>
<td>Contour Map of Long-run Barge Rate Effects Using Kernel Smoothing Based on Land Distances and Water Distances (SD=50)</td>
<td>32</td>
</tr>
</tbody>
</table>
Chapter 1

Grain Transport on the Mississippi River and Spatial Corn Basis

1.1 Introduction

1.1.1 Corn Production and Exports

Corn is the most widely produced feed grain in the United States, accounting for more than 90 percent of total value and production of feed grains. The United States is the largest producer and exporter of corn in the world, with approximately 20 percent of the corn crop exported to other countries.\(^1\)

In 2007, more than 60 percent of U.S. corn was harvested in the Midwest. Demand for corn is geographically diverse, creating areas of deficit throughout the West, the Southeast, and Northeast. Corn demanded by overseas markets is exported through ports in the Gulf, the Pacific Northwest, the Atlantic Coast, and the Great Lakes. Figure 1.1 demonstrates how this imbalance of surplus and deficit creates the need for long distance transportation.

Figure 1.1: Corn surplus/deficit map with the transportation system
1.1.2 Corn Transportation

Corn is transported to distant markets in two patterns - one for domestic use and the other for export. Trucks supply most of the transportation for the domestic market, and barges and rail supply the export market. From 2000 to 2006, trucks supplied, on average, about 68% of the corn used by the domestic market (Figure 1.2). During the same period, barges transported 64% of corn exports. Rail handled about 33% of the export market and 30% of the domestic market. Barges are the main mode of transportation for corn moving to port regions for export.

<table>
<thead>
<tr>
<th>Year &amp; Type of Movement</th>
<th>CORN</th>
<th>Rail</th>
<th>Barge</th>
<th>Truck</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1,000 Tons</td>
<td>Percent</td>
<td>1,000 Tons</td>
<td>Percent</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>68,984</td>
<td>30%</td>
<td>37,831</td>
<td>16%</td>
</tr>
<tr>
<td>2001</td>
<td>75,833</td>
<td>31%</td>
<td>38,864</td>
<td>16%</td>
</tr>
<tr>
<td>2002</td>
<td>72,615</td>
<td>31%</td>
<td>41,598</td>
<td>18%</td>
</tr>
<tr>
<td>2003</td>
<td>71,443</td>
<td>30%</td>
<td>36,488</td>
<td>15%</td>
</tr>
<tr>
<td>2004</td>
<td>77,377</td>
<td>32%</td>
<td>37,302</td>
<td>15%</td>
</tr>
<tr>
<td>2005</td>
<td>77,908</td>
<td>30%</td>
<td>31,739</td>
<td>12%</td>
</tr>
<tr>
<td>2006</td>
<td>91,552</td>
<td>32%</td>
<td>34,587</td>
<td>12%</td>
</tr>
<tr>
<td>Average</td>
<td>76,216</td>
<td>31%</td>
<td>36,916</td>
<td>15%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year &amp; Type of Movement</th>
<th>EXPORT</th>
<th>CORN</th>
<th>Rail</th>
<th>Barge</th>
<th>Truck</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1,000 Tons</td>
<td>Percent</td>
<td>1,000 Tons</td>
<td>Percent</td>
<td>1,000 Tons</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>15,213</td>
<td>29%</td>
<td>35,150</td>
<td>66%</td>
<td>2,594</td>
</tr>
<tr>
<td>2001</td>
<td>15,822</td>
<td>30%</td>
<td>36,904</td>
<td>68%</td>
<td>1,806</td>
</tr>
<tr>
<td>2002</td>
<td>14,527</td>
<td>27%</td>
<td>38,125</td>
<td>79%</td>
<td>Not available</td>
</tr>
<tr>
<td>2003</td>
<td>14,371</td>
<td>30%</td>
<td>32,872</td>
<td>69%</td>
<td>364</td>
</tr>
<tr>
<td>2004</td>
<td>17,422</td>
<td>33%</td>
<td>33,974</td>
<td>64%</td>
<td>1,978</td>
</tr>
<tr>
<td>2005</td>
<td>20,251</td>
<td>40%</td>
<td>28,778</td>
<td>57%</td>
<td>1,600</td>
</tr>
<tr>
<td>2006</td>
<td>28,145</td>
<td>44%</td>
<td>31,941</td>
<td>50%</td>
<td>3,342</td>
</tr>
<tr>
<td>Average</td>
<td>17,936</td>
<td>33%</td>
<td>33,821</td>
<td>64%</td>
<td>1,598</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year &amp; Type of Movement</th>
<th>DOMESTIC</th>
<th>CORN</th>
<th>Rail</th>
<th>Barge</th>
<th>Truck</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1,000 Tons</td>
<td>Percent</td>
<td>1,000 Tons</td>
<td>Percent</td>
<td>1,000 Tons</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>53,771</td>
<td>30%</td>
<td>2,681</td>
<td>2%</td>
<td>119,938</td>
</tr>
<tr>
<td>2001</td>
<td>57,911</td>
<td>31%</td>
<td>2,960</td>
<td>2%</td>
<td>124,034</td>
</tr>
<tr>
<td>2002</td>
<td>58,288</td>
<td>32%</td>
<td>3,473</td>
<td>2%</td>
<td>119,835</td>
</tr>
<tr>
<td>2003</td>
<td>57,072</td>
<td>30%</td>
<td>3,656</td>
<td>2%</td>
<td>127,552</td>
</tr>
<tr>
<td>2004</td>
<td>59,955</td>
<td>32%</td>
<td>3,328</td>
<td>2%</td>
<td>124,511</td>
</tr>
<tr>
<td>2005</td>
<td>57,657</td>
<td>28%</td>
<td>2,961</td>
<td>1%</td>
<td>148,918</td>
</tr>
<tr>
<td>2006</td>
<td>63,407</td>
<td>29%</td>
<td>2,646</td>
<td>1%</td>
<td>155,744</td>
</tr>
<tr>
<td>Average</td>
<td>58,280</td>
<td>30%</td>
<td>3,095</td>
<td>2%</td>
<td>131,519</td>
</tr>
</tbody>
</table>

Figure 1.2: Corn modal shares

Millions of goods are moved on the inland waterways by barges. Figure 1.3 shows 2007 barge tonnage by major commodity group on U.S. inland waterways. The Mississippi

---

2Source: USDA report: Study of Rural Transportation Issues
The river and its attributes are agriculturally significant waterways (Figure 1.4). The majority of the traffic flows on the Mississippi, Illinois and Ohio rivers. On the lower Mississippi river, food and farm products are major commodities shipped, which take account of about 80% of total annual flows of water borne commodities (Figure 1.5). The importance of the Mississippi River water system largely results from an efficient barge transportation system that links the north central United States and the Cornbelt to the Gulf. Most corn exports are shipped through the Mississippi Gulf region - 63% of all corn volume exported in 2007. In transporting grains, barges and railroads compete with each other for long haulage. Trucking is a complement to barges and railroads in transporting grains from producers or inland markets to elevators along the river or railway stations for long distance shipment.

![2007 Barge Tonnage, by Major Commodity Group](image)

Figure 1.3: 2007 barge tonnage, by major commodity group

Interregional trade theory predicts that in a competitive market price differences between markets are explained by transportation cost. The precise role played by transportation costs can depend on distances that grain needs to travel and also on the extent
Figure 1.4: Agriculturally-significant waterways

Figure 1.5: Agricultural and total freight moving on U.S. waterways
to which markets are integrated into the transportation system. The economic problem posed by corn prices variation along with transport is to what extent transportation costs affect prices, and how the effects are related to geographical locations and characteristics of markets. In this paper, I derive an equilibrium model of how transportation costs affect the spatial pattern of corn prices and test its implications and find that the effect of barge rate changes on the Mississippi are geographically diverse.

A Theory of Spatial Basis

In this paper, I define the spatial basis of corn at a market as the corn price in the market minus the contemporaneous corn price at New Orleans, Louisiana (NOLA). In competitive markets, spatial basis in equilibrium should be explained by transportation costs.

Figure 1.6 illustrates a market for grain (market i) and the destinations to which it ships. An ethanol plant is located to the west of i, at which the equilibrium price of corn is $P_E$. To the east of i lies the export market. Grain traveling that route is shipped overland to a river market, where the price is $P_R$. At the river market, grain is loaded onto barges and shipped down the Mississippi to New Orleans, where the price is $P_{NO}$. From New Orleans the grain is shipped overseas.

I consider the allocation of a fixed amount of grain, $Q_T$, harvested near i and to be shipped from there. If grain travels to both inland and export destinations from market i, the price of grain purchased at i for export must equal the price of grain purchased at i for use at the ethanol plant. Further the price at i must account for the transport costs west to the ethanol plant, as well as east to the river market and south to New Orleans. These equilibrium conditions can be represented as:

$$P_E(Q_E) - c_E(Q_E) = P_R - c_R(Q_R)$$  \hspace{1cm} (1.1) \\
$$P_R = P_{NO} - bD_R$$  \hspace{1cm} (1.2) \\
$$Q_T = Q_E + Q_R,$$  \hspace{1cm} (1.3)

where $Q_E$ is the quantity of grain shipped from market i to the ethanol plant, $c_E(Q_E)$ is the cost of shipping grain from market i to the ethanol plant, $Q_R$ is the quantity of grain shipped from market i to the river, $c_R(Q_R)$ is the cost of shipping grain from market i to the river, b is the cost per bushel per mile of shipping grain by barge down the
Mississippi, and $D_R$ is the river distance from the river market to New Orleans.

In the expressions above $P_{NO}$ - the export price - is assumed to be unaffected by variations in grain flows from market $i$, while $P_E$ - the price paid by the ethanol plant - is allowed to vary with $Q_E$: $P_E$ is a non-increasing function of $Q_E$. The two overland unit transport cost functions, $c_E$ and $c_R$, are allowed to increase with quantities shipped, reflecting the optimization of shippers who are presumed to use cheaper means of transport before more expensive means.

Consider the comparative statics of the system resulting from an exogenous change in $b$, the cost of barge transport along the Mississippi. First, substitute from equations (1.2) and (1.3) into (1.1) to obtain:

$$P_E(Q_T - Q_R) - c_E(Q_T - Q_R) = P_{NO} - bD_R - c_R(Q_R).$$  \(1.4\)

Totally differentiating (1.4) with respect to $b$ and $Q_R$ results in:

$$-P_E'dQ_R + c'_EdQ_R = -D_Rdb - c'RdQ_R.$$  \(1.5\)
Equation 1.5 can be solved for the equilibrium response in $Q_R$ to a change in the cost of barge shipping:

$$\frac{dQ_R}{db} = \frac{D_R}{P'_E - c'_E - c'_R} < 0.$$  \hfill (1.6)

The change in basis at market $i$ can be deduced from (1.6) and the following equilibrium relation derived from equations (1.1)-(1.3):

$$P_i - P_{NO} = -bD_R - c_R(Q_R).$$  \hfill (1.7)

Equations (1.6) and (1.7) together imply:

$$\frac{d(P_i - P_{NO})}{db} = -D_R \left( \frac{P'_E - c'_E}{P'_E - c'_E - c'_R} \right).$$  \hfill (1.8)

Further noting that $P'_E - c'_E$ is the slope of the inverse demand at market $i$ for grain for shipment to the ethanol plant and that $-c'_R$ is the slope of the inverse demand at market $i$ for grain for export via the river market, expression (1.8) can be written in terms of the direct elasticities of those demand curves and the shares of market $i$ grain going to each:

$$\frac{d(P_i - P_{NO})}{db} = -D_R \left( \frac{1}{1 + \frac{\eta_E \alpha_E}{\eta_R \alpha_R}} \right) < 0,$$  \hfill (1.9)

where $\alpha_E$ and $\alpha_R$ are the shares of $Q_T$ shipped to the ethanol plant and the river, and $\eta_E$ and $\eta_R$ are the direct price elasticities of demand, observed at market $i$, for the two uses of grain. Both elasticities are defined to be negative (non-positive) numbers. Equation (1.9) yields:

**Empirical prediction 1:** Basis at market $i$ declines with an exogenous increase in the barge rate.

Note some limiting cases of equation (1.9), the equilibrium response of market-$i$ basis to a change in barge rates. First, if $\alpha_E = 0$ and market $i$ ships only to the river, the right-hand side of (1.9) is equal to $-D_R$. The effect at market $i$ of a change in barge rates is the same at $i$ as it is at the river market. Further, if we consider (1.9) as describing markets along the river at different distances from New Orleans (and for which $\alpha_E = 0$), the magnitude of the basis response to barge rates is proportional to river distance $D_R$. This implies:
**Empirical prediction 2:** The response of basis to barge rate changes for markets on the river is greater the higher up the river.

Also note from (1.9) that the effect of a very elastic river-destination demand is the same as that of a low $\alpha_E$: if the river-destination demand is highly elastic and serves to nearly fix the price at market $i$ to equal the transportation-cost-adjusted price at the river, then the right-hand-side of (1.9) is equal to $-D_R$ and the effect of a barge rate change inland is the same as that felt at the river. Because the form of the comparative static result in (1.9) involves the ratio $\eta_E/\eta_R$, the effect of a barge rate change will be larger in magnitude the more elastic river-destination demand is relative to inland-destination demand. Similarly, the share ratio $(\alpha_E/\alpha_R)$ effect says that the effect of a barge rate change will be larger in magnitude the smaller is the share of market that is sent inland.

Both the elasticity ratio and share ratio effects are plausibly related to the distance between market $i$ and the river, $D_{iR}$. For greater distances from the river, markets are more likely to be shipping larger shares of their grain to destinations other than the river; and for those destinations, the elasticity of demand from the inland destination is more likely to be small. For both reasons, I posit a third empirical implication of the model:

$$
\frac{d(P_i - P_{NO})}{db} = -D_R f(D_{iR}), \quad \text{where} \quad 0 < f(D_{iR}) < 1 \quad \text{and} \quad f'(D_{iR}) < 0. \quad (1.10)
$$

In words, I have:

**Empirical prediction 3:** The magnitude of the barge-rate effect on basis declines with distance from the market to the river.

To restate the second and third empirical predictions: (2) for markets on the river (or shipping all of their grain to the Mississippi), $f(D_{iR}) = 1$ and the response of basis to barge rate changes is proportional to $D_R$, the distance upriver from New Orleans; (3) for markets away from the river, the response of basis to barge rates is smaller than that at the river, and the magnitude of the effect declines with distance to the river. Further, based on equation (1.9), for markets $i$ and $j$ with the same distance to the river, $D_{iR} = D_{jR}$, if $\frac{\eta_i}{\eta_R} \frac{\alpha_i}{\alpha_R} > \frac{\eta_j}{\eta_R} \frac{\alpha_j}{\alpha_R}$, market $i$ has smaller barge-rate effect when $D^i_R = D^j_R$, where $D^i_R$ and $D^j_R$ are the river distances for market $i$ and $j$, respectively. Even if market $i$ is farther up the river, $D^i_R = \alpha D^j_R$ with $\alpha > 1$, if $\frac{\eta_i}{\eta_R} \frac{\alpha_i}{\alpha_R} >> \frac{\eta_j}{\eta_R} \frac{\alpha_j}{\alpha_R}$, the barge-rate effect...
on market i can be smaller than on market j. This implies that market i is less integrated into the river transportation system. It may be because a large share of demand for corn at market i comes from an ethanol plant and a smaller share comes from the river market for export. This logic implies:

**Empirical prediction 4:** Markets that are more integrated into the river transportation system tends to exhibit more pronounced barge-rate effects.

I test these predictions and measure the sizes of the associated effects with a mixture of parametric and nonparametric methods.

### 1.2 Data

More than 3.5 million observations from an unbalanced panel of daily corn prices in over 4,000 markets from 2005 to 2010 are matched with daily observations on barge rates, quoted as cents per bushel from St. Louis to New Orleans\(^3\) and diesel prices\(^4\), quoted as cents per gallon. Figure 1.8 plots the universe of markets for which daily corn prices are available. To focus on markets most likely to be integrated into the river transport system, I filtered out markets that are farther than 150 miles from the river system. A market is included in the resulting data set if its straight-line distance to the nearest of the major rivers (Figure 1.7) is less than 150 miles and if there are more than 45 non-missing observations\(^5\). Because the stretch of rivers north of McGregor in Minnesota\(^6\) are frozen from December to March every year, markets whose latitude greater than 46.606667 are excluded. There are 1,189 markets that remain, which are shown in Figure 1.9.

---

\(^3\)Daily corn prices and barge rates data are provided by GeoGrain.

\(^4\)Diesel price is weekly data from the USDA website. The days on which observations are missing were filled with the latest diesel price that is available.

\(^5\)The minimum of 45 observations was chosen to ensure that degrees of freedom in the regression are greater than 30.

\(^6\)Coordinates: 46.606667, -93.313889.
Figure 1.7: The Rivers

Figure 1.8: All Markets
Figure 1.9: Corn Markets within 150 miles of major navigable rivers
1.3 An Econometric Model of Corn Basis and Transportation Cost

The time series regression model specifies basis in market \( i \) on day \( t \) as a linear function of barge rates and diesel prices on days \( t, t-1, \) and \( t-2 \), as well as a periodic seasonal term:

\[ b_{i,t} = \alpha_i + \beta_{i,0} p_{t - i}^b + \beta_{i,1} p_{t - 1}^b + \beta_{i,2} p_{t - 2}^b + \gamma_{i,0} p_t^d + \psi_{i,t} + \varepsilon_{i,t}, \tag{1.11} \]

where \( b_{i,t} \) is corn spatial basis in market \( i \) on day \( t \); \( p_{t - i}^b \) are St. Louis-to-NOLA barge rates on day \( t, t-1 \) and \( t-2 \) for \( i = \{0, 1, 2\} \), in cents/bushel; and \( p_t^d \) is contemporaneous diesel price in cents/galon. \( \psi_{i,t} \) comprises four pairs of Fourier components:

\[ \psi_{i,t} = \sum_{s=1}^{4} \left( \lambda_s \cos \frac{2\pi st}{365} + \phi_s \sin \frac{2\pi st}{365} \right). \]

Diesel prices are used to proxy for prices of other transportation modes, such as trucks and railroads. Corn basis also exhibits seasonality across years. Therefore, a Fourier component, which repeats annually is included in the model to capture the seasonal movement of basis. The question of how many Fourier terms are to be included is an open one. Adding higher frequency terms allows more rapid change during the year.

The regression model in which the choice of numbers of lagged barge rates is carried out by two systematic procedures. The first procedure, called forward selection in this context, starts with a model composed of contemporaneous and one lagged barge rates, contemporaneous diesel price, constant terms and four pairs of Fourier components. I test the addition of each lagged barge rate using a chosen model comparison criterion and add one more lagged barge rate that improves the model until the next one does not. Tables 1.1 and 1.2 show the test results. The proportion of regressions with a significant F-test based on 5% or 10% levels starts decreasing at \( p = 3 \) and increases by a small percent when \( p = 4 \). When \( p = 5 \), 85% of regressions have significant barge rate effects. The proportion of regressions in the joint tests of barge rates increases by less than 20% from \( p = 1 \) to \( p = 5 \). Although there is not a clear cut of number lagged barge rates to be included in the model shown from forward selection method, the result shows the significance of the contemporaneous and the first-lagged barge rates in most of regressions. Since forward selection method tends to under fit the data, backward selection is also considered in the model selection procedure.
Table 1.1: Forward Selection: Proportion of regressions that are significant at 10% level

<table>
<thead>
<tr>
<th>Models</th>
<th>T-tests</th>
<th>F-tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>p=1, q=0</td>
<td>$\alpha = 0$ $\beta_0 = 0$ $\beta_1 = 0$</td>
<td>$\gamma_0 = 0$ $\beta_0 = \beta_1 = 0$</td>
</tr>
<tr>
<td></td>
<td>0.8982 0.5837 0.4154</td>
<td>0.7373 0.7102</td>
</tr>
<tr>
<td>p=2, q=0</td>
<td>$\alpha = 0$ $\beta_0 = 0$ $\beta_1 = 0$ $\beta_2 = 0$</td>
<td>$\gamma_0 = 0$ $\beta_0 = \ldots = \beta_2 = 0$</td>
</tr>
<tr>
<td></td>
<td>0.8992 0.7286 0.2101 0.2782</td>
<td>0.7017 0.7655</td>
</tr>
<tr>
<td>p=3, q=0</td>
<td>$\alpha = 0$ $\beta_0 = 0$ $\beta_1 = 0$ $\beta_2 = 0$ $\beta_3 = 0$</td>
<td>$\gamma_0 = 0$ $\beta_0 = \ldots = \beta_3 = 0$</td>
</tr>
<tr>
<td></td>
<td>0.8907 0.6558 0.1862 0.1529 0.2664</td>
<td>0.6729 0.7481</td>
</tr>
<tr>
<td>p=4, q=0</td>
<td>$\alpha = 0$ $\beta_0 = 0$ $\beta_1 = 0$ $\beta_2 = 0$ $\beta_3 = 0$ $\beta_4 = 0$</td>
<td>$\gamma_0 = 0$ $\beta_0 = \ldots = \beta_4 = 0$</td>
</tr>
<tr>
<td></td>
<td>0.8768 0.5898 0.1743 0.1752 0.1303 0.3592</td>
<td>0.6417 0.7711</td>
</tr>
<tr>
<td>p=5, q=0</td>
<td>$\alpha = 0$ $\beta_0 = 0$ $\beta_1 = 0$ $\beta_2 = 0$ $\beta_3 = 0$ $\beta_4 = 0$ $\beta_5 = 0$</td>
<td>$\gamma_0 = 0$ $\beta_0 = \ldots = \beta_5 = 0$</td>
</tr>
<tr>
<td></td>
<td>0.8597 0.6503 0.2049 0.1403 0.1439 0.1248 0.4690 0.4754 0.8525</td>
<td></td>
</tr>
</tbody>
</table>

Note: $p =$ number of lagged barge rates. $q =$ number of lagged of diesel prices.
### Table 1.2: Forward Selection: Proportion of regressions that are significant at 5% level

<table>
<thead>
<tr>
<th>Models</th>
<th>T-tests</th>
<th>F-tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>p=1, q=0</td>
<td>$\alpha = 0$ $\beta_0 = 0$ $\beta_1 = 0$</td>
<td>$\gamma_0 = 0$ $\beta_0 = \beta_1 = 0$</td>
</tr>
<tr>
<td></td>
<td>0.8801 0.5008 0.3456</td>
<td>0.6814 0.6494</td>
</tr>
<tr>
<td>p=2, q=0</td>
<td>$\alpha = 0$ $\beta_0 = 0$ $\beta_1 = 0$ $\beta_2 = 0$</td>
<td>$\gamma_0 = 0$ $\beta_0 = \ldots = \beta_2 = 0$</td>
</tr>
<tr>
<td></td>
<td>0.8790 0.6445 0.1378 0.2000</td>
<td>0.6303 0.7126</td>
</tr>
<tr>
<td>p=3, q=0</td>
<td>$\alpha = 0$ $\beta_0 = 0$ $\beta_1 = 0$ $\beta_2 = 0$ $\beta_3 = 0$</td>
<td>$\gamma_0 = 0$ $\beta_0 = \ldots = \beta_3 = 0$</td>
</tr>
<tr>
<td></td>
<td>0.8736 0.5406 0.1435 0.1025 0.1802</td>
<td>0.5995 0.6926</td>
</tr>
<tr>
<td>p=4, q=0</td>
<td>$\alpha = 0$ $\beta_0 = 0$ $\beta_1 = 0$ $\beta_2 = 0$ $\beta_3 = 0$ $\beta_4 = 0$</td>
<td>$\gamma_0 = 0$ $\beta_0 = \ldots = \beta_4 = 0$</td>
</tr>
<tr>
<td></td>
<td>0.8512 0.5088 0.1197 0.1153 0.0924 0.2650</td>
<td>0.5792 0.7077</td>
</tr>
<tr>
<td>p=5, q=0</td>
<td>$\alpha = 0$ $\beta_0 = 0$ $\beta_1 = 0$ $\beta_2 = 0$ $\beta_3 = 0$ $\beta_4 = 0$ $\beta_5 = 0$</td>
<td>$\gamma_0 = 0$ $\beta_0 = \ldots = \beta_5 = 0$</td>
</tr>
<tr>
<td></td>
<td>0.8251 0.5683 0.1211 0.0829 0.0774 0.0701 0.4035</td>
<td>0.4144 0.8051</td>
</tr>
</tbody>
</table>

Note: $p =$ number of lagged barge rates. $q =$ number of lagged of diesel prices.
The backward selection starts with a model composed of all candidate lagged barge rates, which are contemporaneous and maximum number of lagged barge rates, which is seven. Proportions of significant regressions of joint tests on subsets of barge rate coefficients are compared. Tables 1.3 and 1.4 show the t-tests and F-tests at 10% and 5% significance levels, respectively. T-tests on each coefficients illustrate that contemporaneous, 1-lagged and 2-lagged barge rates are significant in 63%, 31%, and 33% of regressions, which are far more than other lagged barge rates\(^7\). Besides, joint tests on the subsets of coefficients of barge rates show that there are more than 50% of regressions are significant in testing the null hypothesis of \(\beta_2 = \ldots = \beta_7 = 0\) at the 10% significance level and about 50% of regressions are significant in testing the same null at the 5% level. Combining the forward and backward selection results to balance the under- and over-fit tendencies of the two approaches, two lagged barge rates are chosen. Therefore, the model contains a contemporaneous, once-lagged and twice-lagged barge rates, seasonality and other control variables.

\(^7\)\(\beta_7\) is exceptional. However, \(\beta_7\) is not selected if lagged barge rates before are not chosen.
Table 1.3: Backward Selection: Proportion of regressions that are significant at 10% level

<table>
<thead>
<tr>
<th>Model:</th>
<th>p=7, q=0</th>
<th>T-tests</th>
<th>F-tests</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\alpha = 0$</td>
<td>$\beta_0 = 0$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.8184</td>
<td>0.6297</td>
</tr>
</tbody>
</table>

Note: p = number of lagged barge rates. q = number of lagged of diesel prices.
Table 1.4: Backward Selection: Proportion of regressions that are significant at 5% level

<table>
<thead>
<tr>
<th>Model: p=7, q=0</th>
<th>T-tests</th>
<th>F-tests</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>α  = 0</td>
<td>β₄  = 0</td>
</tr>
<tr>
<td></td>
<td>β₀  = 0</td>
<td>β₁  = 0</td>
</tr>
<tr>
<td></td>
<td>0.7874</td>
<td>0.2246</td>
</tr>
<tr>
<td></td>
<td>β₂  = 0</td>
<td>β₃  = 0</td>
</tr>
<tr>
<td></td>
<td>β₃  = 0</td>
<td>β₄  = 0</td>
</tr>
<tr>
<td></td>
<td>β₄  = 0</td>
<td>β₅  = 0</td>
</tr>
<tr>
<td></td>
<td>β₅  = 0</td>
<td>β₆  = 0</td>
</tr>
<tr>
<td></td>
<td>β₆  = 0</td>
<td>β₇  = 0</td>
</tr>
<tr>
<td></td>
<td>β₇  = 0</td>
<td>γ₀  = 0</td>
</tr>
<tr>
<td></td>
<td>β₀  = 0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>β₂  = 0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>β₃  = 0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>β₄  = 0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>β₅  = 0</td>
<td></td>
</tr>
</tbody>
</table>

Note: p = number of lagged barge rates. q = number of lagged of diesel prices.
All variables in model (1.11) are quasi-differenced to account for serial correlation, as OLS estimates show serial correlation in the residuals. Therefore, an estimate of \( \hat{\rho} \) from an OLS regression on the residual model \( \varepsilon_{i,t} = \rho \varepsilon_{i,t-1} + u_{i,t} \) is used to transform variables:

\[
\begin{align*}
b_{i,t}^* &= b_{i,t} - \hat{\rho} b_{i,t-1}, \\
X_t^* &= X_t - \hat{\rho} X_{t-1}, \\
\alpha_i^* &= \alpha_i (1 - \hat{\rho}), \text{and} \\
\varepsilon_{i,t}^* &= \varepsilon_{i,t} - \hat{\rho} \varepsilon_{i,t-1},
\end{align*}
\]

where \( X_t^* = [p_t^b, p_{t-1}^b, p_{t-2}^b, p_{i,t}^d] \) for all \( t \geq 3 \).

The sum of barge-rate effects in three time periods \( t \), \( t-1 \) and \( t-2 \) accounts for "long-run" barge-rate effects. The interpretation of the long-run effect is if, for example, barge rates change and are then sustained for a long period (three days in the model), the ultimate effect on spatial basis will be \( \beta_{i,0} + \beta_{i,1} + \beta_{i,2} \) times the barge rate change.

## 1.4 Estimates

### 1.4.1 Regression Results

**Seasonal Components**

Of the 1,189 markets that satisfy the distance-to-the-river and degrees-of-freedom criteria, 83% have seasonal components significant at the 10% level. Across the estimates, the median difference between seasonal peak and seasonal trough is approximately 7 cents/bushel. Figure 1.10 displays the estimated seasonal factors for five markets whose seasonal ranges were close to this median value. Their locations are displayed on the map in Figure 1.11.

Although there are variations in seasonal pattern for the five displayed markets, seasonal components tend to peak in April, May or June and troughs occur around August, September and October.
Figure 1.10: Estimated Seasonal Factor for Markets #2551, #1423, #1014, #4796 and #4092

Figure 1.11: Markets #4796, #2551, #4092, #1423 and #1014 from North to South
Barge Rate Coefficients

The coefficient $\beta_0$ reflects the contemporaneous effect of barge rate on basis. Table 1.5 is the summary of tests on three barge rate coefficients.

T-tests from the 1,189 regressions show that 72% of the $\beta_0$'s are significant and 99% of significant ones are negative. Fewer $\beta_1$ and $\beta_2$ estimates are significant, most of which are negative (1.5). The results imply that barge rate effects are largely reflected in corn prices contemporaneously. One-day and two-day lagged barge rates have mild effects and the effects are negative. From the distribution of estimated barge rate effects shown in Figures 1.12, 1.13 and 1.14, contemporaneous and lagged barge rates effects are negative, which is consistent with our empirical prediction 1 that basis declines with an exogenous increase in the barge rate.
Table 1.5: Statistics and tests on barge rates coefficients

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>% of significance</th>
<th>% of negative among significant ones</th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>72%</td>
<td>99%</td>
<td>-0.25</td>
<td>-0.23</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>20%</td>
<td>74%</td>
<td>-0.04</td>
<td>-0.02</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>29%</td>
<td>87%</td>
<td>-0.04</td>
<td>-0.05</td>
</tr>
</tbody>
</table>

F-tests

<table>
<thead>
<tr>
<th>$H_0$: $\beta_0 = \beta_1 = \beta_2 = 0$</th>
<th>$H_0$: $\beta_0 + \beta_1 + \beta_2 = 0$</th>
<th>Mean of sum</th>
<th>Median of sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>77%</td>
<td>78%</td>
<td>-0.34</td>
<td>-0.32</td>
</tr>
</tbody>
</table>
F-tests show that 77% of the $\beta_i$'s are jointly significant, which verifies that barge-rate changes explain part of changes in spatial basis. To test the long-run (three-day) effect of barge rates, F-tests on the sum of all $\beta_i$'s equal to zero, for each market, indicates that the sum is significant in 78% of regressions. Figure 1.15 shows the distribution of long-run effect. The mean of the sum is -0.34, and the median is -0.32, which means that a one cent/bushel increase in the barge rate at St. Louis from shipping corn to NOLA will induce a 0.34 cents/bushel drop, on average, in corn price in the Midwest if the corn price at NOLA unchanged. The effect is economically significant considering that the price of corn is around 600 cents/bushel and barge rates are volatile, with range of changes around 80 cents/bushel.
Figure 1.13: Histogram of $\beta_1$

Figure 1.14: Histogram of $\beta_2$
Diesel Price Coefficients

As to diesel price, across markets 15% of the estimates of $\gamma_0$ are significant at the 10% level, and 75% of these are negative. Compared with barge rates, diesel prices are less statistically significant in explaining corn prices since fewer regressions show significant diesel price, which can be due to the collinearity between barge rates and diesel prices and also due to the fact that daily diesel prices are imputed from weekly observations.

1.4.2 Smoothing and Mapping

Nonparametric smoothing techniques provide an exploratory graphical tool to uncover the behavior of the estimates. The visual information on barge rate and diesel price effects on basis in markets can be shown by applying those methods. Kernel smoothing and robust locally weighted regression techniques are applied to long-run barge rate and long-run diesel price effects to estimate a smoothed surface so that the spatial pattern of barge rate effects and diesel price effects can be visually observed.
A kernel method is applied to smooth effects. The surface is smoothed based on a 60-by-60 grid evenly spaced between the minimum and maximum of latitude and longitude of observations. The length of a degree of latitude is approximately 69 miles. A degree of longitude varies in size. At the equator, it is approximately the same size as a degree of latitude, 69 miles. The size gradually decreases to zero as the meridians converge at the poles. For most of our data around 45 degree of latitude, a degree of longitude is approximately 53 miles. Our data span approximately 14 degrees in latitude and 17 degrees in longitude. Therefore, the lengths of every grid for 60-by-60 grids are approximately 16 miles in latitude and 15 miles in longitude.

Data points that do not fall on the grids are “moved” to their closest point on the grid. The estimate at each grid point is a weighted average of all observations. Weights are determined by the bivariate normal distribution centered at the estimating grid. So means of two independent variables (latitude and longitude) are zeros. The kernel smooth assumes zero covariance and equal standard deviations. The standard deviation controls the smoothness of the surface. Larger standard deviations will create smoother surfaces.
but the trade-off is more bias.

Figures 1.17 and 1.18 show contour plots of smoothed long-run barge rate effects on spatial basis in 1,189 markets using bivariate normal kernel smoothing with standard deviation equal to 50. Figure 1.18 is identical to Figure 1.17 but also shows market locations. Contour lines are at -0.5, -0.4, -0.2, -0.1 and 0. Contour lines stretch out along the Mississippi River, which implies that markets that are closer to the Mississippi River have more negative effects. As markets get farther away to the west and the east of the Mississippi river, the effects eventually fade out. It is consistent with empirical prediction 3. It can be noted from the shading area of the contour plots that barge rate effects are not symmetric with respect to the Mississippi River. Markets west of the Mississippi River have barge rate effects fading out more rapidly with respect to their distance to the river. This makes sense because there is little barge traffic on the Missouri river. The Illinois and Ohio rivers to the east of the Mississippi are two major transportation waterways. As noted, the most negative effects are around St. Louis, which are smaller than -0.5. St. Louis is the port with the most capacity for loading/unloading grains on the Mississippi river waterway system. The majority of corn in St. Louis is loaded for shipment to the Gulf. This implies that $\alpha_E/\alpha_R$ in equation (1.9) is small and that the magnitude of the barge-rate effect is big at St. Louis. This also supports the empirical prediction 4 that the markets that are largely integrated into the river transportation system exhibit more pronounced barge-rate effects.

Figures 1.19 and 1.20 show contour plots with different standard deviations of 30 and 70. The contour lines for Figure 1.19 are -0.6, -0.4, -0.2 and 0. The contour maps present similar spatial patterns of barge rate effects on basis with different degrees of smoothness.

With regard to the contemporaneous diesel price, no spatial pattern is shown\(^8\). Consistent with the histogram of diesel price effect shown before, diesel price effects are near zero for most markets. As noted earlier, it may be due to multicollinearity.

**Locally Weighted Regression**

Univariate locally weighted regression, or loess, which was first discussed by Cleveland (1979), is a method for smoothing a scatterplot, $(x_i, y_i), i = 1, \ldots, n$, in which the fitted value at $x_k$ is the value of a polynomial fit to the data using weighted least squares,\(^8\) The contour plot is omitted.
Figure 1.17: Contour Map of Barge Rate Effects Using Kernel Smoothing (SD = 50 miles)

Figure 1.18: Contour Map of Barge Rate Effects Using Kernel Smoothing (SD = 50 miles)
Figure 1.19: Contour Map of Barge Rate Effects Using Kernel Smoothing (SD = 30 miles)

Figure 1.20: Contour Map of Barge Rate Effects Using Kernel Smoothing (SD = 70 miles)
where the weight for \((x_i, y_i)\) is large if \(x_i\) is close to \(x_k\) and small if it is not. A robust fitting procedure is used that guards against deviant points distorting the smoothed points. A multivariate smoothing procedure is an extension of univariate locally weighted regression. In my context, an estimate of the regression surface at any value \(x\) in the 2-dimensional space of the independent variables, which are latitude and longitude, is the value of a polynomial fit of the independent variables in a neighborhood. Each point in the neighborhood is weighted according to its Euclidean distance from \(x\).

Contour plots of smoothed barge rate effects on spatial basis in 1,189 markets based on bivariate locally weighted regression smoothing are shown in Figures 1.21 and 1.22. Figure 1.22 is identical to Figure 1.21 but also shows market locations. The proportion of data used in local regression is 0.2 and it is linear regression (degree = 1). Contour lines are at -0.6, -0.4, -0.2 and 0. Similar spatial pattern is shown using LOESS. Similar LOESS smoothing with the same parameter selections are applied to diesel price effects. No particular spatial pattern is observed.

Figure 1.21: Contour Map of Barge Rate Effects Using loess Smoother

\[\text{Figure is omitted.}\]

---

\(9\) Figure is omitted.
Figure 1.22: Contour Map of Barge Rate Effects Using loess Smoother w/ Markets

Kernel Smoothing of Long-run Barge Rate Effects with Respect to Land Distances and Water Distances

To examine how a market’s relative location to the rivers can affect barge rate effects in a different perspective, kernel smoothing are applied to the long-run barge rate effects with respect to land distances and water distances. Land distance is the straight-line shortest distance of a market to a river. Markets that are east of the Mississippi River are denoted with negative distances and those that are west of the river are positive. Water distance is the waterway distance of the market to the port at New Orleans, MO. Figures 1.23, 1.24 and 1.25 are the contour maps using kernel smoothing with different standard deviations. Figure 1.23 shows that for the same level of water distance, barge rate effects are smaller when absolute value of land distances is bigger. Barge rate effects are asymmetric between markets west and east of the Mississippi River. Barge rate effects die out slowly for markets east of the river. Markets that are most sensitive to barge rates are located around St. Louis.
Figure 1.23: Contour Map of Long-run Barge Rate Effects Using Kernel Smoothing Based on Land Distances and Water Distances (SD=50)
Figure 1.24: Contour Map of Long-run Barge Rate Effects Using Kernel Smoothing Based on Land Distances and Water Distances (SD=30)
Figure 1.25: Contour Map of Long-run Barge Rate Effects Using Kernel Smoothing Based on Land Distances and Water Distances (SD=70)
1.5 Concluding Remarks

I study how changes in transportation cost affect prices in domestic corn markets within 150 miles of the Mississippi River transportation system. Specifically, I focus on effects of changes in barge rates and in diesel prices on spatial basis. I run unrestricted linear regressions using feasible generalized least square (FGLS) for each market based on daily data from 2005 to 2011. Of 1,189 regressions, the majority of them show that three-day barge rates are both individually and jointly significant at the 10% level, and less so contemporaneous diesel price. The long-run effects of barge rates, which are the sums of three-day coefficients, are significant for over 70% of regressions. The mean is -0.34: if the barge rate were to increase by 1 cent per bushel, the average corn price in the markets relative to the corn price in the Gulf would fall 0.34 cents per bushel after three days. The long-run effects of diesel price are also significant for most regressions, and the average effect is -0.06.

To obtain a surface that shows how markets are integrated into the river transportation system, I apply nonparametric smoothing algorithms – kernel and loess – to the long-run coefficients of barge rates and contemporaneous diesel price. Contour maps show that corn prices in markets, which are close to St. Louis and close to the Mississippi River and Illinois River, are more negatively affected by barge rates. The effects are less pronounced in markets that are farther away from the rivers, which shows the importance of the Mississippi waterway system in transporting corn for export. Those markets, such as markets around St. Louis, which has large proportion of grains loaded onto barges on the rivers for export, tend to show more negative effects. In contrast, for markets in the Midwest, which has more demanding regions beside the Gulf of Mexico, such as PNW and other markets in Texas, the barge-rate effects on basis are smaller compared to that in St. Louis. Generally, markets that are greatly integrated into the river system have large magnitude of effects. With respect to diesel prices, the effects of diesel prices are much smaller, which are almost zero. There are no particular spatial patterns present in the smoothed plots, which may due to multicollinearity between barge rates and diesel prices and the weekly diesel price data used.

This paper analyzes the economic problem of to what extent transportation costs affect prices, and how the effects are related to geographical locations and characteristics of markets. Understanding what drives changes in the spatial patterns of grain prices is critical to those engaged in grain production, transport, and trade. Ongoing calls for
improving the logistical efficiency and hydrologic health of the Mississippi River have attracted attention from policy makers, agricultural producers and consumer groups, and the transportation industry. It is important for policy makers to understand how the river transportation system can influence the agricultural economy and it is important for grain market participants at a more micro level. To my knowledge, there has been no such work investigating the spatial pattern of barge rate effects on grain prices.

One of the limitations of this paper is the lack of daily rates of other transportation modes in the model. Future work may incorporate more frequent diesel price observations or truck and rail rates. Furthermore, barge rate effects on spatial pattern of prices may be dynamic since corn and transportation markets have experienced big changes over years, for example, the ethanol mandate in 2005 dramatically increases domestic demand of corn. Therefore, time effects may be considered in future work.
Chapter 2

The Information Content of Futures Prices in China’s Soybean Futures Market

2.1 Introduction

A commodity futures price is often viewed as a forecast of a future spot price. (See, for example, Cootner (1960), Dusak (1973), Breeden (1980) and Hazuka (1984).) The price at $t$ of a contract for delivery at $t+1$ is an unconditionally unbiased forecast of future spot price at $t+1$ if:

$$F_{t+1}^t = E(S_{t+1}),$$  \hspace{1cm} (2.1)

where $F_{t+1}^t$ is the $t$-dated price of a $t + 1$-delivery contract and $S_{t+1}$ is the spot price at time $t + 1$. This suggests that without using other information, the futures price $F_{t+1}^t$ itself is an unbiased estimate of future spot price. To empirically test unconditional unbiasedness, if we assume a linear relationship between spot and futures prices, we can write:

$$S_{t+1} = a + bF_{t+1}^t + u_{t+1},$$  \hspace{1cm} (2.2)

where $u_{t+1}$ is a stationary error in forecasting future spot price $S_{t+1}$ and $E(u_{t+1}) = 0$. Risk premium is zero means if $a = 0$, and $b = 1$, $F_{t+1}^t$ is an unconditional unbiased estimate of $S_{t+1}$.

Besides the current futures price $F_{t+1}^t$, past futures prices and spot prices are also
available and may contain information about the future spot price. In an efficient futures market, all information contained in past futures prices and past spot prices ought to be embedded in the current futures price $F_{t+1}$. If not, current futures price $F_{t+1}$ as a forecast of future spot price $S_{t+1}$ can be improved in the sense of reducing bias by adding past information. We say that $F_{t+1}$ is a conditionally unbiased forecast of $S_{t+1}$ if

$$F_{t+1} = E(S_{t+1}|\Omega_t),$$

(2.3)

where $\Omega_t$ denotes price information available at time $t$. An empirical model is:

$$S_{t+1} = a^c + b^c F_{t+1} + u_{t+1}.$$  

(2.4)

Conditional unbiasedness means that $a^c = 0$, $b^c = 1$ and $E(u_{t+1}|\Omega_t) = 0$.

The risk premiums $a$ in equation (2.2) or $a^c$ in equation (2.4) are zero when market participants are risk neutral. When risk premiums are positive, the market is said to be in normal backwardation. Futures prices will rise over the life of the contract. It happens when short hedgers, such as producers, sell futures contracts at a price below the expected future spot price to avoid price risk. The difference between the two prices, the risk premium, compensates purchasers of futures contracts for bearing the spot price risk. Conversely, long hedgers may be willing to buy futures contracts at a price above the expected future spot price to avoid price risk implying negative risk premia and contango: price of futures contracts tending to fall over time. Whichever type of hedger dominates in terms of trading volume over a particular time period will determine whether the risk effect is positive or negative.

This paper tests unconditional (i.e. $a = 0$ and $b = 0$ in equation 2.2) and conditional (i.e., $a^c = 0$, $b^c = 0$ and $E(u_{t+1}|\Omega_t) = 0$ in equation 2.4) unbiasedness of futures price $F_{t+1}$ as a forecast of future spot price $S_{t+1}$ under two frameworks. Approaches in the two frameworks are different because of assumptions about the existence of unit roots in price series. Besides futures price, other price variables are considered as alternatives in forecasting future spot price. This paper also compares the forecasting ability of $F_{t+1}$ with other forecast variables based on certain criterion.
2.2 Literature Review

There exist a large number of literatures in testing unbiasedness and efficiency of futures prices as forecasts of future spot prices. An efficient market is one which reflects all information that is available in the market. Competitive conditions will force prices to adjust instantaneously to a new price in which new information is reflected in it. Fama (1970) classified the efficiency into three forms, the weak form, semi-strong form and strong form based on meanings of “all information.” The weak form of efficiency is when the information reflected by the market is just historical prices; semi-strong form extends further to include all public available information; and strong form reflects any information available including insider information. Empirical testing for efficiency is difficult because the definition is so general. Empirical work on the efficiency typically focuses on the adjustment of futures prices to a particular information set, for example information contained in past futures and spot prices.

Some of the early tests for efficiency assume that prices follow a martingale stochastic process. A martingale is a stochastic sequence of variables and its major property is that the conditional expected value of the random variable at time $t + 1$ equals the value at time $t$, i.e., price at $t + 1$ incorporates information contained in past prices. However, Danthine (1977) and Lucas (1978) have theoretically shown that periodic failure of the Margingale property to hold does not necessarily imply market inefficiency. Danthine criticizes Samuelson’s (1965) argument and develop reasons why linkage between martingale and market efficiency is problematic.

The finding of no direct or causal relationship between martingale property of futures prices and market efficiency places emphasis on the forecasting ability of futures prices. Rausser and Carter (1983) examine the efficiency of the U.S. soybean complex, including soybean, soybean oil, and soybean meal futures markets by investigating their forecasting ability in terms of both bias and variability measures via structurally based ARIMA models. They conclude that the constructed models significantly “outperform” the futures market for soybeans and soybean meal but not soybean oil for both “long- and short-range forecasts” based on the mean-square prediction error criterion.

Early works on testing forecasting ability of futures prices include Gray and Tomek (1970), Kofi (1973) and Leuthold (1974). They investigate forecast accuracy of futures prices for storable and/or non-storable commodities. Gray and Tomek find that the intertemporal price relationships between pre-harvest quoted futures prices for harvest
time contract and spot prices at harvest time differ between storable and non-storable commodities. According to their results, futures prices are unbiased forecasts for corn and soybeans, while it is not for potatoes. Kofi supports Gray and Tomek’s findings that continuous inventory futures markets such as wheat, corn and soybeans outperform the discontinuous inventory futures markets such as potato and cocoa markets based on $r^2$ performance tests. Both Kofi and Leuthold find that futures markets are unbiased forecasts of spot prices for near-maturity dates only. Their models are estimated by regressions of two price series via ordinary least squares. Various versions of the fundamental model, a regression of future spot prices on its forecasts futures prices can be seen to test various hypothesis in commodity markets, for example Martin and Garcia (1981) tested whether forecasting performance of futures markets for live cattle and hogs changes over time and whether differs with cyclical variation in prices and seasonally. For more versions of fundamental model of testing forecasting ability of futures prices in live cattle or live hog markets, see Kolb and Gray (1983), Hayenga et al. (1984), and Kastens and Schroeder (1995).

Ma (1989) examines the forecasting accuracy of energy futures markets for crude oil, heating oil, and leaded gasoline during the period of 1980 to 1986. The approach is to take first-differences of the variables and re-estimate the forecasting regressions. Ma compares the forecasting accuracy of futures prices with forecasts generated from other time series models: random walk, univariate and multivariate ARIMA, unrestricted and Bayesian VAR. Ma found that, on average, futures markets outperform these econometric models for one-, two-, three-, and six-month out-of-sample forecasts based on the criterion of the root mean squared error. Zulauf, Irwin, Ropp and Sberna (1996) evaluate the forecasting ability of the December corn futures contract and November soybean futures contract during spring 1995. They estimate a regression equation of cash price changes on the futures-cash basis suggested by Fama and French (1987). Their results show that the spring-time quotes of the harvest futures are unbiased estimates of prices at harvest.

After development of cointegration techniques by Engle and Granger (1987), cointegration techniques have been widely used in testing market efficiency and unbiasedness of futures prices. When time series are nonstationary, conventional t- and F-tests on regression coefficients are not appropriate since residuals are not normally distributed. Instead, cointegration techniques have been applied by Hakkio and Rush (1989), Baillie and Bollerslev (1989), and Barnhart and Szakmary (1991) to test unbiasedness in foreign exchange markets. Hakkio and Rush uses a result of theory of cointegration that two
prices from a pair of efficient markets cannot be cointegrated to test for efficiency in the German and United Kingdom foreign exchange markets. They fail to reject the joint hypothesis of no risk premium combined with efficient use of information for both markets. Baillie and Bollersleve verify the existence of a unit root or a stochastic trend in daily spot and forward exchange rates by univariate test and the existence of unit roots in a set of jointly determined exchange rates by multivariate tests. They conclude with evidence for the presence of a unit root for seven currencies’ spot and forward exchange rates tested, which are currencies of UK, West Germany, France, Italy, Switzerland, Japan and Canada, and they are cointegrated. Further, their result on the set of currencies suggests that the exchange rates are bounded together by strong trend components. Barnhart and Szakmary have the same conclusion that both spot and forward rates for the U.K., Germany, Japan, and Canada have unit roots and are cointegrated. However, they also find that unbiasedness of forward rates for all four currencies is rejected via an alternative error correction specification.

Baillie and Myers (1991), Schroeder and Goodwin (1991), Quan (1992) and Schwartz and Sakmary (1994) have found mixed evidence of cointegration between cash and futures prices for storable commodities, but no cointegration for nonstorable commodities. Other researchers (Brenner & Kroner, 1995; Zapata and Fortenbery, 1996) have suggested that the empirical finding of no cointegration for storable commodities could be due to a misspecification problem, that is, the exclusion of possible nonstationary elements of the cost of carry, particularly stochastic interest rates in the cointegration system. Beck (1994) and McKenzie and Holt (2002) apply the cointegration techniques to agricultural commodity futures markets. They estimate an error correction model that takes into account the cointegration relationship between the futures and future spot prices. Beck tests the unbiasedness and efficiency of five commodity futures markets, cattle, orange juice, corn, copper and cocoa at eight and twenty-four week horizon and find mixed results on market efficiency for five commodities with different forecast horizons. McKenzie and Holt test market unbiasedness and efficiency in five commodity futures markets - live cattle, hogs, corn, and soybean meal - using cointegration and error correction models with GARCH-in-mean processes. They conclude that each market is unbiased in the “long run”, and cattle, hogs and corn futures markets are “short-run” inefficient and biased.

Many empirical works (more can been found from Bessler and Covey (1991), Schroeder and Goodwin (1991), Fortenbery and Zapata (1993, 1997), and Sabuhoro and Larue
(1997)) tested the unbiasedness hypothesis on the basis of cointegration. One of necessary conditions for cointegration technique to be valid is the assumption that time series are random walks. However, the question of whether commodity prices are random walks or trend stationary has never been settled. Several studies in testing if commodity futures prices are trends or random walks provide controversial results. Larson (1960) found evidence to support that prices move randomly, which means price changes reflect new information and hence approximate a random variation. Stevenson and Bear (1970) concluded that corn and soybean futures prices move in a systematic, as opposed to a random manner. Several recent papers discuss the question at a general level. Blough (1992a, b), Cochrane (1991), and Sims (1989) argued that the question of whether a time series has a unit root is inherently unanswerable on the basis of a finite sample of observations.

There are few studies on the efficiency of commodity markets in China. Williams et al. (1998) studied mung bean trading and found that conditions for arbitrage existed. Wang and Ke (2005) examined the efficiency of Chinese wheat and soybean futures markets prior to 2002 and their results suggest a long-term equilibrium relationship between the futures price and cash price for soybeans and weak short-term efficiency in the soybean futures market. They found the futures market for wheat to be inefficient.

2.3 Methodology

If spot and futures prices are stationary, equation (2.2) can be estimated by OLS. The estimates exhibit the standard distribution properties under normal assumptions. To test unconditional unbiasedness of \( F_{t+1} \), conventional t and F tests can be applied to test if \( a = 0 \) and \( b = 1 \) in equation (2.2) when there is no serial correlation in residuals. If \( u_{t+1} \) is serially correlated, it means that past futures and/or spot prices are relevant in forecasting future spot price, thus \( F_{t+1} \) is conditionally biased. In such circumstance, one may choose to explicitly include the past futures and spot prices in the model. Specifically, suppose that \( u_{t+1} \) follows

\[
    u_{t+1} = \rho u_t + v_{t+1},
\]

(2.5)
where $v_{t+1}$ is white noise. Substitute into (2.2) from (2.5) to get:

$$S_{t+1} = a(1 - \rho) + bF_t^{t+1} + \rho S_t - \rho b F_{t-1}^t + v_{t+1}. \quad (2.6)$$

Equation (2.6) can be estimated by non-linear least squares with restrictions on the relationship among coefficients. A general unrestricted version of equation (2.6) is

$$S_{t+1} = a' + b'F_{t+1}^t + c'S_t + d'F_{t-1}^t + v'_{t+1} \quad (2.7)$$

and can also be estimated. One can test the significance of current and past futures prices and past spot prices to examine if past information is useful in forecasting future spot price.

If spot and futures prices are both nonstationary, estimates by OLS still are consistent, but conventional tests can no longer be used. In this case, the cointegration framework provides a general mathematical form such that both unconditional and conditional unbiasedness of $F_{t+1}^t$ can be tested with restrictions imposed on the model. In addition to nonstationarity, price series should have other properties to be suited for a cointegration model. Specifically, if both series require first differencing to render each series stationary, i.e., they are integrated of order one, and some linear combination of the two prices is stationary, then futures and spot prices are cointegrated. To estimate $b$ in equation (2.2) is to estimate the cointegrating relationship between $F_{t+1}^t$ and $S_{t+1}$.

Cointegrated series can be rewritten in an error correction representation as described in Granger (1986). To write the relationship in error correction form, the cointegrated time series should be contemporaneous; that is, the two prices should be observable at the same time. However, the economic relationship of interest is between two non-contemporaneous time series, $F_{t+1}^t$, which is available at time $t$, and $S_{t+1}$, whose value is revealed at $t+1$. This mis-matching problem can be solved by incorporating properties of time series. Specifically, one can obtain a cointegrating relationship between two contemporaneous prices series, $S_{t+1}$ and $F_{t+1}^{t+2}$. If $S_{t+1}$ and $F_{t+1}^{t+2}$ are cointegrated, then $S_{t+1}$ and $F_{t+1}^{t+1}$ are cointegrated as well. The reason for this is that futures price is I(1), which is a necessary condition for cointegration.

$$F_{t+1}^{t+2} = c + F_{t}^{t+1} + w_{t+1}, \quad (2.8)$$

where $w_{t+1}$ is stationary with zero mean. If $F_{t+1}^{t+2} \overset{\text{coint}}{\sim} S_{t+1}$, then $c + F_{t}^{t+1} + w_{t+1} \overset{\text{coint}}{\sim} S_{t+1}$,
which is equivalent to $F_{t}^{t+1 \text{ cont}} \sim S_{t+1}$ because $c$ and $w_{t+1}$ do not change the cointegrating relationship. Specifically, if the cointegrating relationship between $F_{t+1}^{t+2}$ and $S_{t+1}$ is

$$S_{t+1} = \theta + \phi F_{t+1}^{t+2} + \eta_{t+1}, \quad (2.9)$$

substitute (2.2) into (2.9), to get:

$$S_{t+1} = \theta + c\phi + \phi F_{t+1}^{t+1} + (\phi w_{t+1} + \eta_{t+1}), \quad (2.10)$$

where $\phi = b$, $\theta + cb = a$ and $\phi w_{t+1} + \eta_{t+1}$ is white noise. Therefore, $a$ and $b$ in equation (2.2) can be recovered from equations (2.8) and (2.9). Thus the cointegrating relationship $a$ and $b$ between $F_{t}^{t+1}$ and $S_{t+1}$ are represented by the cointegrating relationship between two contemporaneous time series $F_{t+1}^{t+2}$ and $S_{t+1}$. The terms $\theta + c\phi$ and $\phi$ can be estimated in the error correction form and $c$ can be recovered from equation (2.8). Thus, the intercept of equation (2.2) $a$ can be recovered. Inferences on $a = 0$ and $b = 1$ in (2.2) for unconditional unbiasedness can also be conducted from the cointegrating relationship in (2.10) using the MLE approach by Johansen and Juselius (1990). In Beck (1994) and McKenzie and Holt (2002), they also intend to use cointegration framework to estimate and test cointegrating relationship between $F_{t}^{t+1}$ and $S_{t+1}$, however the time periods of cointegrated time series in the error correction representation are misaligned, which leads to different and possibly incorrect estimates and test results.

Following Granger (1986), if $F_{t+1}^{t+2}$ and $S_{t+1}$ are cointegrated as shown in equation (2.9), then

$$\Delta S_{t+1} = \theta + \varphi (S_{t} - bF_{t}^{t+1}) + \sum_{i=0}^{m} \beta_{i} \Delta F_{t-i}^{t+1} + \sum_{j=0}^{n} \gamma_{j} \Delta S_{t-j} + \nu_{t+1}, \quad (2.11)$$

Note that equation (2.11) is derived from the cointegrating relationship between $S_{t+1}$ and $F_{t+1}^{t+1}$, but contains the relationship of original economic interest between $S_{t+1}$ and $F_{t}^{t+1}$. The transformed series are now stationary so coefficient estimates are asymptotically normally distributed and conventional t- and F-tests on hypothesis are valid in testing conditional unbiasedness. The residual $\nu_{t}$ is stationary with zero mean.

Conditional unbiasedness implies the additional restrictions: $\varphi = -1$ and $\beta_{i} = \gamma_{j} = 0$ for all $i$ and $j$. To illustrate the restrictions for conditional unbiasedness, Equation (2.11)
is rewritten below by adding $S_t$ to both sides:

$$S_{t+1} = \theta + (1 + \varphi)S_t - \varphi b F_{t+1}^t + \sum_{i=0}^{m} \beta_i \Delta F_{t-i+1}^t + \sum_{j=0}^{n} \gamma_j \Delta S_{t-j} + \nu_{t+1},$$

(2.12)

The coefficients of lagged spot and futures price changes, $\beta_i$ and $\gamma_j$, are zero because past information is already completely incorporated in the current futures price. If $\varphi = -1$ does not hold, then past spot price contribute information useful for predicting $S_{t+1}$, therefore all available information is not fully reflected in the current futures price $F_{t+1}^t$. For the same reason, conditional unbiasedness also implies that $\nu_{t+1}$ is serially uncorrelated. Another necessary condition for conditional unbiasedness is $\theta = 0$ in equation (2.12). Testings of $\theta = 0$ and $b = 1$ can be performed directly from error correction representation of equation (2.11). Estimate of $a$ and hypothesis test of $a = 0$ are implied from equation $a = \theta + bc$, where $c$ can be estimated from (2.8).

The error correction model derived from the cointegrating relationship transformed the nonstationarity in the time series to a general model with stationary time series, from which conditional unbiasedness can be tested based on conventional t- and F-tests. Besides error correction model, other stationary-inducing transformations of Equation (2.2) have been applied before and after the cointegration technique is developed. But they turn out to be special cases of equation (2.11) with different constraints. For example, stationary in equation (2.2) can be achieved by differencing both $S_{t+1}$ and $F_t$:

$$\Delta S_{t+1} = A + B \Delta F_{t+1}^t$$

(2.13)

In terms of equation (2.11), restrictions $\varphi = 0$ and $\beta_i = \gamma_j = 0$ for $i \neq 0$ and all $j$ are imposed. Equation (2.13) is misspecified because $\varphi \neq 0$ for non-stationary cointegrated variables.

Another variation of equation (2.2) regresses spot price changes at $t+1$ on the futures basis at time $t$:

$$\Delta S_{t+1} = A + B(F_{t+1}^t - S_t).$$

(2.14)

Equation (2.14) is a special case of (2.11) with some constraints for testing unconditional and conditional unbiasedness in equation (2.11) imposed. Equation (2.14) can be obtained by imposing $b = 1$ and $\beta_i = \gamma_j = 0$ for all $i$ and $j$ in (2.11). Tests of $A = 0$ and $B = 1$ actually test the unbiasedness hypothesis but impose $\beta_i = \gamma_j = 0$. The general
model (2.11) is preferred for testing conditional and unconditional unbiasedness because all restrictions, including $\beta_i = \gamma_j = 0$ are tested.

### 2.4 Data

Unconditional and conditional unbiasedness tests are applied to China’s No.1 soybean market. The Dalian Commodity Exchange (DCE) is the only futures market for soybean trading. Dalian is located in the northeast of China and has convenient access to ocean, rail and road transportation facilities. It is an important trading and shipping hub in the main soybean production regions in China. No.1 soybean contracts create non-GMO (non-genetically modified) soybean trading market. Unlike GMO soybeans, most of which are imported and mainly traded as No.2 soybeans at DCE, non-GMO soybeans are grown in China. Similar to the U.S. soybean market, most of the non-GMO soybeans in China are harvested from September through October. There are six contracts each year for No.1 soybeans: January, March, May, July, September and November.

Since $S_{t+1}$ is the spot price at contract expiration, the frequency of observations (the number of times per year that the futures market generate a forecast) is limited to the number of contracts offered per year. Furthermore, futures price must be chosen at a forecast horizon less than or equal to the observation interval to avoid introducing residual correlation by overlapping observation intervals (Granger and Newbold, 1977, p.115). Two-month-ahead futures contracts are most actively traded in the market. Thus by choosing the futures price two months prior to expiration, all six expiration dates can be pooled. Futures price for contracts from March 2003 to March 2013 are available. Therefore there are 61 observations. Two-month-ahead futures price is selected as the futures price on the first Thursday of the month two month prior to expiration. Spot prices at the futures expiration date are represented by the first Wednesday futures price of the expiring contract. Theoretically, spot and futures prices are the same at expiration since arbitrage will drive them together. The use of futures price data avoids biases introduced by inaccurate spot price data.

Figure 2.1 shows the futures price of expiring contracts in delivery month, which is approximated by the first Wednesday price of the month. The plot is also an approximate of spot prices in the same month. The range of spot prices over this period is from as low as about 2500 RMB per metric ton to over 5500 RMB per metric ton. Table 2.1 shows...
the descriptive statistics of the data.

2.5 Estimation

To address the forecasting power of futures prices in a comprehensive way, models are developed to test the unbiasedness of futures price under two assumptions: time series are trend-stationary (or stationary); and time series are random walks. With the trend stationary assumption, the usual OLS $t$ and $F$ statistics, calculated in the usual way, have the same asymptotic distributions as they do for stationary regressions. Table (3.5) shows the OLS regression results of model (2.2). T tests on $a = 0$ and $b = 1$ are insignificant at the 5% level. The joint test is also insignificant at the same significance level. If we accept the hypothesis of $a = 0$ and $b = 1$, then the test results support the statement that $F_t^{t+1}$ is an unconditionally unbiased forecast of $S_{t+1}$. A more restrictive hypothesis that $F_t^{t+1}$ is also conditionally unbiased can be tested by checking the coefficient of autoregression of residuals. If the coefficient is high, it indicates there is strong serial
Table 2.1: Descriptive statistics of futures prices of expiring contracts

<table>
<thead>
<tr>
<th>Unit</th>
<th>RMB / metric ton</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Obs.</td>
<td>61</td>
</tr>
<tr>
<td>Mean</td>
<td>3600</td>
</tr>
<tr>
<td>Std</td>
<td>814.25</td>
</tr>
<tr>
<td>Median</td>
<td>3643</td>
</tr>
<tr>
<td>Min</td>
<td>2409</td>
</tr>
<tr>
<td>Max</td>
<td>5714</td>
</tr>
</tbody>
</table>

Table 2.2: Estimated OLS Regression for Model (2.2)

Model (2.2): \[ S_{t+1} = a + bF_t^{t+1} + u_{t+1} \]

<table>
<thead>
<tr>
<th>(a)</th>
<th>(b)</th>
<th>(H_0 : b = 1) p-value: 0.07</th>
<th>(H_0 : a = 0 \text{ and } b = 1) p-value: 0.18</th>
</tr>
</thead>
<tbody>
<tr>
<td>307.41*</td>
<td>0.91***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(182.22)</td>
<td>(0.05)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Significant at 10% significance level. ** Significant at 1% significance level.
correlation in the residual. The estimate of $\rho$ in equation (2.5) is 0.1936, which is fairly small; therefore not much serial correlation exists. No serial correlation implies that past spot and futures prices do not contain information useful in forecasting $S_{t+1}$. From this perspective, futures price is also conditionally unbiased.

Although there is no strong autocorrelation in the residual based on estimate of $\rho$, past futures and spot prices are imposed in the model to estimate a different model and examine how each price series explain variations in future spot price. Table (2.3) shows the results for equation (2.7). Hypothesis testing results show that only the intercept and the last period spot price $S_t$ are significant at the 10% significance level. The t-test of $S_t = 1$ is insignificant at the 5% level. This regression provides evidence to the statement that price series are random walks. A random walk model may be considered to forecast future spot price, i.e.,

$$S_{t+1} = S_t + \epsilon_{t+1}, \quad (2.15)$$

where $\epsilon_{t+1}$ is stationary.

<table>
<thead>
<tr>
<th>Table 2.3: Estimation of Model (2.7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model (2.7): $S_{t+1} = a' + b'F_{t+1} + c'F_{t-1} + d'S_t + \nu_{t+1}$</td>
</tr>
<tr>
<td>$a'$</td>
</tr>
<tr>
<td>356.32*</td>
</tr>
<tr>
<td>(186.38)</td>
</tr>
<tr>
<td>DW: 2.01</td>
</tr>
</tbody>
</table>

*** Significant at the 1% level. * Significant at the 10% level.

Under the assumption that the time series are unit root processes, the estimation and tests of restrictions are to be conducted under the cointegration framework if variables satisfy certain time series properties. The usual tests to examine if data are possibly unit
root processes are the augmented Dicky-Fuller and Phillips-Perron test. However, Dickey and Pantula (1987) methodology should be applied first to test that at most one unit root for each price series. The augmented Dickey-Fuller (1981) test is based on the initial assumption of a single unit root. Dickey and Pantula (1987) showed that the augmented Dickey-Fuller (1981) test can yield incorrect conclusions if more than one root actually exists. Dickey-Pantula test suggests sequentially testing for three unit roots, two unit roots, then one unit root. The three steps procedure starts regressing $\Delta^3 Y_t$ on $\Delta^2 Y_{t-1}$, then on $\Delta^2 Y_{t-1}$ and $\Delta Y_{t-1}$, then on $\Delta^2 Y_{t-1}$, $\Delta Y_{t-1}$ and $Y_{t-1}$, where $\Delta^d$ indicates of the degree of differencing, and $Y_t$ represents the time series. The null hypothesis for the first step is that the coefficients of $\Delta^2 Y_{t-1}$ is zero, i.e., there are three unit roots against the alternative hypothesis that there are two unit roots. The null hypothesis for step two and three are that the coefficients of $\Delta Y_{t-1}$ and $Y_{t-1}$ are zero, respectively. The procedure does not stop testing until it fails to reject the null hypothesis. Test statistics are compared to the tables of Fuller (1976). In table 2.4, the Dickey-Pantula tests show that there is, at most, a single unit root in each series.

Table 2.4: Dickey-Pantula tests

<table>
<thead>
<tr>
<th>Price Series</th>
<th>$DP_1$</th>
<th>$DP_2$</th>
<th>$DP_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_{t+1}$</td>
<td>$-12.82^{***}$</td>
<td>$-3.88^{***}$</td>
<td>$-2.31$</td>
</tr>
<tr>
<td>$S_{t+1}$</td>
<td>$-11.78^{***}$</td>
<td>$-3.94^{***}$</td>
<td>$-2.27$</td>
</tr>
</tbody>
</table>

*** The null hypothesis was rejected at 1% significance level. Test statistics are compared to the Fuller (1976) table.

Note: $DP_i$ tests the null hypotheses $\alpha_i = 0$ with $\alpha_j$, where $j > i$ constrained to zero.

Although it is difficult to conclude whether a time series is difference stationary for finite observations, the augmented Dickey-Fuller test and Phillips-Perron test are applied to test if time series have one unit root and also are difference stationary instead of trend stationary. The optimal number of augmenting lags for the model in the augmented
Dickey-Fuller test is determined by using Akaike’s information criterion. The AIC criterion suggests two lags for the futures price series. Table 2.5 shows the statistics of the ADF tests based on the regression model with and without trend. T statistic, $T_ρ$, of the test of the null hypothesis that there is one unit root, i.e., $ρ = 0$, is -3.39. The test fails to reject the null at the 5% significance level. The joint test on the null hypothesis that there is no trend and there is a unit root is rejected at the 5% significance level. The t statistic of a test that $ρ = 0$ based on a model without trend is -2.31, which fails to reject at the 10% level. The joint test that the intercept is zero and there is a unit root fails to reject at the 10% level. Based on a 5% significance level, the ADF tests show that there is a unit root and no trend in the futures price series. The null of a unit root process is not rejected based on the results of the Phillips-Perron test shown in Table 2.6.

Since spot price is approximated by the futures price, we expect that the tests on the spot price will give the same conclusions. Test results on the spot price are omitted.

Phillips-Ouliaris tests for cointegration between $S_{t+1}$ and $F_{t+1}^{t+2}$ is conducted for Equation (2.10) and with the two variables reversed because the cointegration technique does not specify which should be the left-hand variable. Both regressions reject the null of no
Table 2.6: The Phillips-Perron Unit Root Test for 2-month-ahead Futures Price

<table>
<thead>
<tr>
<th>Type</th>
<th>Lags</th>
<th>$\rho$</th>
<th>$Pr &lt; \rho$</th>
<th>$\tau$</th>
<th>$Pr &lt; \tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero Mean</td>
<td>1</td>
<td>0.1778</td>
<td>0.7203</td>
<td>0.2327</td>
<td>0.7503</td>
</tr>
<tr>
<td>Single Mean</td>
<td>1</td>
<td>-6.0886</td>
<td>0.3233</td>
<td>-1.7805</td>
<td>0.3864</td>
</tr>
<tr>
<td>Trend</td>
<td>1</td>
<td>-10.6287</td>
<td>0.3561</td>
<td>-2.3308</td>
<td>0.4105</td>
</tr>
</tbody>
</table>

cointegration at a 1% significance level\(^1\).

To test unconditional unbiasedness under the cointegration framework, Johansen’s cointegration procedure is applied to estimate $\theta$ in Equation (2.9). Test results in Table 2.7 shows that we fail to reject $b = 1$ at the 5% significance level, but we reject the joint test of $\theta = 0$ and $b = 1$ at the 5% level\(^2\). Furthermore, since $a = \theta + bc$, we can estimate $a$ if we get estimates of $b$ and $c$. Differencing futures time series from equation (2.15) and estimating $c$ by OLS, we get $\hat{c} = 35.28$ and the t-statistic is 0.81. Thus we fail to reject the null of $c = 0$ at the 1% significance level. If we take $c = 0$, we get $\hat{a} = \hat{\theta} = 12.46$, which is statistically significant at the 5% significance level. Therefore, we can conclude that without relying on past spot or futures prices, current futures price $F_{t+1}$ is a biased estimate of future spot price $S_{t+1}$. The bias is positive, coming from the risk premium. It means that the No.1 soybean futures market is in normal backwardation. Farmers or market participants who sell futures contracts are willing to sell soybeans at about 12.46 RMB per metric ton less than the expected future spot price in order to hedge price risk.

To estimate $\theta$ and $b$ when considering past information, the error correction model (2.11) is estimated by OLS since the transformed series are stationary. The results are shown in Table 2.9. The model was estimated with one lag of $\triangle S_t$ and $\triangle F_{t+1}$ (which is equivalent to including two lags in the VAR level model). The term $S_t - b F_{t+1}$ was recovered from the cointegrating regression equation (2.9) with $\phi = 1$ to estimate Equation (2.11). Residual serial correlation was not detected by Lagrange multiplier statistics, which were

\(^1\)Results omitted.
\(^2\)Remember that $b = \phi$. 

52
Table 2.7: Unconditional Unbiasedness Tests under Cointegration

<table>
<thead>
<tr>
<th>Model (2.9):</th>
<th>( S_{t+1} = \theta + \phi F_{t+1}^2 + \eta_{t+1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta )</td>
<td>12.46</td>
</tr>
<tr>
<td>( \phi )</td>
<td>0.98</td>
</tr>
<tr>
<td>( H_0 : \phi = 1 )</td>
<td>p-value: 0.40</td>
</tr>
<tr>
<td>( H_0 : \theta = 0 ) and ( \phi = 1 )</td>
<td>p-value: 0.02</td>
</tr>
</tbody>
</table>

Table 2.8: Estimates of Random Walk Model 2.8

<table>
<thead>
<tr>
<th>Model (2.8):</th>
<th>( F_{t+1}^2 = c + F_{t+1}^1 + \eta_{t+1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate:</td>
<td>( c )</td>
</tr>
<tr>
<td></td>
<td>35.28</td>
</tr>
<tr>
<td></td>
<td>(43.61)</td>
</tr>
</tbody>
</table>
computed for up to sixth order residual serial correlation.

A Wald test of the null hypothesis of conditional unbiasedness with restrictions of \( \theta = 0 \), \( \varphi = -1 \), and \( \beta_0 = \gamma_0 = 0 \) is not rejected at the 5% significance level. Therefore, the futures price is a conditionally unbiased forecast of future spot price.
Table 2.9: Estimated Error Correction Model

Model (2.11): \[ \Delta S_{t+1} = \theta + \varphi (S_t - b F_{t+1}) + \sum_{i=0}^{m} \beta_i \Delta F_{t-i+1} + \sum_{j=0}^{n} \gamma_j \Delta S_{t-j} + \nu_{t+1}, \text{ where } m = n = 1 \]

<table>
<thead>
<tr>
<th>(\theta)</th>
<th>(\varphi)</th>
<th>(\beta_0)</th>
<th>(\gamma_0)</th>
<th>(H_0: \theta = 0, \varphi = -1, \text{ and } \beta_0 = \gamma_0 = 0)</th>
<th>(LM^b)</th>
<th>(DW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-40.77</td>
<td>-1.36**</td>
<td>-0.17</td>
<td>0.18</td>
<td>p-value: 0.96</td>
<td>p-value: 0.69</td>
<td>1.99</td>
</tr>
<tr>
<td>(54.71)</td>
<td>(0.63)</td>
<td>(0.53)</td>
<td>(0.51)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^b\) is a test with first-order autocorrelation. **Significant at the 5% level.
2.5.1 Out-of-Sample Fits

In this section, fitting errors from five forecast models are compared based on three criteria: Mean Percentage Error (MPE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Percentage Error (RMSPE). Table 2.10 shows their fitting errors in out-of-sample predictions. To overcome the small sample size problem in comparing out-of-sample predictions between five models, I first estimate the model by leaving out one observation and then calculate the predicted error for the observation using the model estimated by the rest of the observations. I apply the procedure to every observation and use the residual from each observation to calculate the three measures.

<table>
<thead>
<tr>
<th>Model</th>
<th>MPE</th>
<th>MAPE</th>
<th>RMSPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) $\hat{S}_{t+1} = F_t$</td>
<td>-0.74%</td>
<td>5.89%</td>
<td>8.19%</td>
</tr>
<tr>
<td>(2) $\hat{S}_{t+1} = a + bF_t$</td>
<td>-0.77%</td>
<td>6.24%</td>
<td>8.48%</td>
</tr>
<tr>
<td>(3) $\hat{S}<em>{t+1} = a' + b' F</em>{t+1} - c' F_{t-1} - d'S_t$</td>
<td>-0.87%</td>
<td>6.76%</td>
<td>8.85%</td>
</tr>
<tr>
<td>(4) $\hat{S}_{t+1} = S_t$</td>
<td>0.73%</td>
<td>6.17%</td>
<td>8.62%</td>
</tr>
<tr>
<td>(5) $\Delta \hat{S}<em>{t+1} = \theta + \varphi (S_t - bF</em>{t+1}) + \beta_0 \Delta F_{t+1} + \gamma_0 \Delta S_t$</td>
<td>-0.45%</td>
<td>6.45%</td>
<td>8.79%</td>
</tr>
</tbody>
</table>

Models (1), (2), (3) and (5) tend to overforecast according to the negative values of MPE. By comparing values for the three criteria, the five models do not present significant differences in forecasting future spot price. Since models (1) and (4) are the simplest models, we may choose to use these two simple models as a forecast of $S_{t+1}$.
Current futures price $F_{t+1}$ or previous period’s spot price $S_t$ are fairly good estimates of future price $S_{t+1}$.

### 2.5.2 Conclusion

The objective of this paper is to examine the information content of futures price as a forecast of future spot price for No.1 soybean market in China. The paper tests unbiasedness of futures price with and without conditioning on past price information. Because of the difficulty of identifying time series as either trend stationary or unit root processes in a finite sample, unbiasedness is tested based on both assumptions. When time series are stationary or trend stationary, OLS and conventional tests are appropriate to test unbiasedness. If time series are unit root processes and cointegrated, cointegration techniques can be applied.

The results indicate that futures prices are unconditionally and conditionally unbiased forecasts of future spot prices under the assumption of stationary (or trend stationary). In the cointegration framework, futures price is unconditionally biased resulting from a positive risk premium, which means that farmer or market participants who sell contracts pay about 12.46 RMB per metric ton more than future spot price to hedge price risks. It is about 0.3% of the future spot price, which is small considering that the mean of the approximated future spot prices is 3600 RMB per metric ton. When conditioning on past information, futures price is unbiased. The OLS regression and the cointegration technique generate the same conclusion with respect to conditional unbiasedness. Both approaches support that futures prices are conditionally unbiased, which means past prices are useless in improving the forecasting ability of futures prices. However, conclusions based on two techniques are controversial for unconditional unbiasedness.

To compare the accuracy of forecasts of future price by several models, three measurements of percentage of forecasting errors in out-of-sample forecasts are computed. The results show that the five models compared do not exhibit significant difference in terms of percentage forecasting errors. Thus simple models can be applied, i.e., only futures price $F_{t+1}$ or last period spot price $S_t$ will produce fairly unbiased forecasts of $S_{t+1}$. 

57
Chapter 3

Water Levels on the Mississippi River and Grain Transport Barge Rates

3.1 Introduction

The Mississippi waterway has been an important transportation channel for shipping grain from the Midwest to the export terminals in New Orleans Louisiana (NOLA). Locks and dams in the Upper Mississippi River are used to facilitate vessels to travel on the river due to poor natural navigation conditions. Rather than narrowing the river and depending solely on water flow, 23 locks and dams were constructed from St. Paul to St. Louis to store water in upstream reservoirs or pools in order to guarantee the depth of 9-foot deep channel for fully-loaded barges. The U.S. Army Corps of Engineers is responsible for managing and maintaining a navigable condition along the Mississippi Waterways. Figure 3.1 shows the Mississippi Rivers. The Lower Mississippi River stretching from Cairo to New Orleans is free flowing. Once the Missouri River joins the Mississippi River at St. Louis, the combined river is wide and deep enough that locks and dams are not needed to maintain a 9-foot deep channel to New Orleans. With no locks slowing river traffic, it takes about 6 days for barges traveling from St. Louis to NOLA\(^1\). However, without locks and dams balancing the water flow in the Lower Mississippi River, maintaining a navigable condition is costly and time-consuming during

\(^1\)See “Shipping Out” by Frank Lessiter.
drought seasons, which has been seen most recently in 2012.

![The Mississippi Rivers](image)

Figure 3.1: The Mississippi Rivers

In general, waterways with drafts of 12 feet or less are considered to be shallow draft\(^2\). In September 2012, water flow in the river south of St. Louis was so low due to the worst drought in 50 years that several grounding incidents happened and the part of river between Cairo and Vicksburg was closed once. At Memphis, Tennessee the river was 7.9' below normal river levels, near the record low of 10.70' below set in 1988\(^3\).

\(^2\)Information is from the COOSA-Alabama River Improvement Association, Inc. [http://www.caria.org/barges_tugboats.html](http://www.caria.org/barges_tugboats.html)

\(^3\)https://www.workboat.com/newslog.aspx?id=18161
According to an article from workboat.com⁴: “the industry was recommending drafts no greater than 9’ in both directions and southbound tows limited to 25 barges, down from 30.”

A typical barge carries 53,571 bushels of corn or 50,000 bushels of wheat or soybeans⁵, which is equivalent to 1,500 short tons of corn, or 1,400 short tons of wheat or soybeans. When water levels are below a certain stage at shallow points in the river, barges cannot be loaded to their full capacities without grounding. Therefore, shipping and operation costs per ton of cargo shipped can be higher when the water level is too low.

The U.S. Army Corps of Engineers has devoted resources to maintaining healthy navigation conditions for vessels by deepening channels and improving access to the rivers. As the importance of the river conditions to commerce and trade it facilitates has become more pronounced due to flood and drought in recent years, questions arise as to the effects of flow variability and Mississippi River levels on the economic benefits the river provides. This paper analyzes an important part of the economic consequences of water level changes - effects on the costs of grain transport, particularly for shipping grains from St. Louis to New Orleans.

3.2 Literature Review

Few papers study the relationship between the Mississippi waterway system and grain barge transportation and commodity prices. The Food and Agricultural Policy Research Institute (FAPRI) at the University of Missouri published a series of reports addressing transportation on the Mississippi Waterways and agricultural markets along the rivers. In that series, Kruse (2004) shows that the geographic patterns of relative prices of corn and soybeans change in grain markets along the Mississippi Waterway system as the upper Mississippi and Illinois Rivers are closed to barges for the season. FAPRI’s reports on transportation capacity in the upper Mississippi and Illinois Rivers and agricultural markets along the rivers emphasize the close relationship between transportation activities and market prices, revenues and flow patterns of grain shipped on the rivers.

In the popular press, news and articles from various sources have reported the impacts of water flow conditions on the transport practices on the rivers, especially during flood

and drought seasons. For example, from workboat.com\(^6\) in September 2012: “…groundings have caused temporary closures, one-way traffic has become common, tow sizes and barge loadings have been decreased...” and from Alan Bjerga from Bloomberg in November 2012: “Mississippi River barge traffic is slowing as the worst drought in five decades combines with a seasonal dry period to push water levels to a near-record low…”

From Karl Plume of Reuters on April 22 2013 on the impacts on Mississippi barge traffic from floods and accidents\(^7\): “Flooding following torrential rains across the central United States forced the U.S. Army Corps of Engineers to close about a dozen locks on the Illinois River and the Mississippi River north of St. Louis late last week...While the conditions are much different than they were this winter, the effects are quite the same. ‘We’re placing operational guidelines on the vessel industry and shutting parts of the river,’ said Coast Guard spokesman Colin Fogarty”

While newspapers and magazines report the impacts of water flows on the transport practice on the Mississippi Rivers, there are no economic studies that measure and evaluate such impacts on barge transportation costs.

### 3.3 Models

How water level changes affect shipping costs can be examined by their effects on barge rates. Changes in barge rates can be influenced by both the demand and supply of barge services. As navigation conditions in the Mississippi affect barge operation practice, the shipping and operation costs in the barge industry change. Higher operation costs in a low water level condition due to serious drought decrease the supply of barges at St. Louis for southbound transport, the effect of which can be transmitted to barge rates, \textit{ceteris paribus}. At the same time, barge rates are driven down as demand decreases due to less demand for barge services from poor grain harvest in a drought season. Other factors, such as the prices of diesel fuel can influence the supply of barges and factors, due to the effect of prices of substitutes or complements for barge services on demand. Seasonality is a factor that influences both demand and supply of barge services for shipping grains from St. Louis to New Orleans. Barges that transport cargo in St. Paul move to the south in winter as the river north of port McGregor, MN freezes for 3 months, which increases

---

\(^6\)https://www.workboat.com/newslog.aspx?id=18161  
\(^7\)http://news.yahoo.com/mississippi-barge-traffic-snarled-floods-accidents-225410069.html
 barges at St. Louis. Heavy traffic on the Upper Mississippi delays transportation, causing barges to be more costly to operate. It conversely decreases barge supply. In harvest season, demand for barges is high and drive barge rates up.

Theoretically, the barge rate \( r \) can be thought of as a function of water level \( w \), diesel price \( p^D \) (representing transportation substitutes and complements), the quantity of barge service supplied (or demanded) \( Q \), seasonality \( s \), and the prices of alternative transportation modes \( p^a \):  

\[
    r = f(w, p^D, Q, s, p^a) \quad (3.1)
\]

Model 3.1 simplifies based on different assumptions about the demand and supply of barge services. As shown in the leftmost graph in Figure 3.2, when barge supply is perfectly elastic, only factors that shift barge supply will influence barge rates. Demand shifts for barge services for example, caused by changes in the availability of alternative transportation modes, or quantity supplied (or demanded) \( Q \) will not affect barge rates. \( Q \) and \( p^a \) are then unnecessary in equation 3.1. When supply is not perfectly elastic, as shown in the rightmost graph in Figure 3.2, changes in barge rates are influenced by a mixture of demand and supply changes, in which case quantity supplied (or demanded) \( Q \) is endogenous and \( p^a \) can be so as well if the cross price elasticity of barge demand is significant. At the other extreme when barge supply is perfectly inelastic, similar arguments can be made and only demand factors influence barge rates. However, no matter if barge supply is perfectly elastic or not, water level \( w \) is predetermined in both cases and transport costs will not affect water levels.

The appropriate estimation approach is different under different assumptions. In the first case, where barge supply is perfectly elastic, water level effects can be estimated by estimating \( r = f(w, s) \) via ordinary least squares (OLS). Water level \( w \) does not correlate with other factors that can affect the barge supply except seasonality, which has been explicitly included in the model. Therefore, OLS estimates of the water level effect are consistent. In the second case where barge supply is not perfectly elastic, quantity supplied \( Q \) is endogenous as might be the alternative transportation mode costs \( p^a \). Thus, only instrumental variables (IV) or full information maximum likelihood (FIML) estimator for the equation’s coefficients will produce consistent estimates. For a detailed discussion and an application of endogeneity testing in supply and demand framework, see Thurman (1986). As the goal of this paper is to estimate water level effects on barge rates, \( p^a \) and \( Q \) can be dropped from the model if water level does not correlate
Figure 3.2: Barge Rate changes by Shifts of Demand and Supply

with them, otherwise the estimate of water level effect is inconsistent. As seen from the estimation results shown in later sections, quantity supplied $Q$ is dropped from the model because its inclusion barely changes the estimate of the water level coefficient. As to the other transportation costs $p^a$, it is assumed that the correlation between water level $w$ and $p^a$ is small, thus $p^a$ is not included in the model for estimating water level effect.
3.4 Data

Daily barge rates for shipping grain from St. Louis to New Orleans, Louisiana (NOLA) from 2005 to 2012 are provided by GeoGrain. Descriptive statistics of barge rates at St. Louis are shown in table 3.1. A plot of barge rates at St. Louis for shipping grain to NOLA is shown in figure 3.4. Daily water level data on the Mississippi River is obtained from the St. Louis district of the U.S. Army Corps of Engineers\(^8\), which collects and maintains daily water level data at certain locations. The Mississippi River at St. Louis (MISL) and Ohio River at Cairo (OHCA) are two key locations where water from the Upper Mississippi River and Ohio River flowing into the Lower Mississippi River are recorded. The coordinates of MISL are 38°37’44” latitude and 90°10’47” longitude. It is 15 miles downstream from the mouth of the Missouri River and at mile 179.6 above the mouth of the Ohio River\(^9\). It is also downstream from the confluence of the Mississippi and Illinois Rivers. OHCA is at river mile 2 upstream from the mouth of Ohio River. Its coordinates are 37°00’00” latitude and 89°30’45” longitude\(^10\). A map showing the river from St. Louis to Memphis with locks and their river miles is shown in figure 3.3\(^11\). Water flows passing through the two locations both influence the water volume in the Lower Mississippi River. Water level is measured by stages. River stages are calculated from an arbitrary “gage zero” point, unique to each location. Zero Gage is 379.94 ft NGVD29 (National Geodetic Vertical Datum of 1929\(^12\)) at MISL, and 270.47 ft NGVD29 at OHCA. At MISL, flood stage is 30 ft. Flood stage is defined as an established gage height for a given location above which a rise in water surface level begins to create a hazard to lives, property, or commerce. The issuance of flood advisories or warnings is linked to flood stage\(^13\). The mean stage since 1861 at MISL is 11.26 ft. During the period of record from 1861, the extreme daily high stage happened in August 1993 at 49.50 ft and the extreme low in January 1940 at -6.20 ft. At OHCA, flood stage is 40 ft and bankfull stage is 44 ft. Bankfull stage is an established gage height at a given location along a river or stream, above which a rise in water surface will cause the river or stream to overflow the lowest

---


\(^10\) According to St. Louis District U.S. Army Corps of Engineers.

\(^11\) Source: [http://www.riverlorian.com/rivermaps.htm](http://www.riverlorian.com/rivermaps.htm)

\(^12\) Sea Level Datum of 1929 was the vertical control datum established for vertical control surveying in the United States of America by the General Adjustment of 1929 - Wikipedia.

natural stream bank somewhere in the corresponding reach\textsuperscript{14}. The mean stage during the period of record since 1858 is 24.24 ft. The extreme daily high of 61.05 ft happened in May 2011, and the extreme daily low of -0.80 ft in December 1871. Descriptive statistics for water level stages at MISL and OHCA are shown in table 3.2.

Figure 3.4 plots the daily water level at MISL from 2005 to January 2013 and barge rates at St. Louis from 2005 to 2012. Figure 3.5 shows the same time series with daily water level replaced by OHCA river stage. Both plots of water levels exhibit seasonality in the series. Water level reaches its peak around April and falls to its lowest around September. Water levels have become more volatile since 2008. The lower water level in 2012 reflects the serious draught in that year. In 2012, water level is at its peak at 20 ft in March, April and May and falls to its lowest at -4.37 ft in December at MISL, which are considerably lower than those for the same periods in other years.


Table 3.1: Descriptive Statistics of Barge Rate at St. Louis

<table>
<thead>
<tr>
<th>Variable</th>
<th>Barge Rate at St. Louis</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Obs.</td>
<td>1,635</td>
</tr>
<tr>
<td>Unit</td>
<td>dollars/short ton</td>
</tr>
<tr>
<td>Mean</td>
<td>14.2267</td>
</tr>
<tr>
<td>Std.</td>
<td>5.4611</td>
</tr>
<tr>
<td>Median</td>
<td>13.2143</td>
</tr>
<tr>
<td>Min.</td>
<td>8.0714</td>
</tr>
<tr>
<td>Max.</td>
<td>43.9286</td>
</tr>
</tbody>
</table>
Table 3.2: Descriptive Statistics of Water Levels at MISL and OHCA

<table>
<thead>
<tr>
<th>Variable</th>
<th>Water Level at MISL</th>
<th>Water Level at OHCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Obs.</td>
<td>2,973</td>
<td>2,574</td>
</tr>
<tr>
<td>Unit</td>
<td>river stage (ft)</td>
<td>river stage (ft)</td>
</tr>
<tr>
<td>Mean</td>
<td>12.0633</td>
<td>24.7486</td>
</tr>
<tr>
<td>Std.</td>
<td>10.3054</td>
<td>11.4418</td>
</tr>
<tr>
<td>Median</td>
<td>9.95</td>
<td>25.41</td>
</tr>
<tr>
<td>Min.</td>
<td>-4.39</td>
<td>7.01</td>
</tr>
<tr>
<td>Max.</td>
<td>38.65</td>
<td>61.05</td>
</tr>
</tbody>
</table>
Figure 3.3: The Mississippi Rivers from St. Louis to Memphis
Figure 3.4: Time Series of Water Level at MISL and Barge Rate at St. Louis
Figure 3.5: Time Series of Water Level at OHCA and Barge Rate at St. Louis
For a quantity measure, the best available data appear to be grain barge movement from St. Louis to New Orleans on day $t$. This is proxied by grain barges unloaded in the New Orleans Region at day $t + 7$, labeled $Q_t$ in the discussion below. The number of barges shipped from St. Louis to NOLA will not be counted as unloading region in the New Orleans until a week later as it usually takes about a week for barges travelling from St. Louis to New Orleans. Weekly barge quantity data is obtained from USDA. Daily data are obtained by dividing weekly barges by seven, the number of days in a week. Thus the weekly barge quantity is transformed into daily data to match frequency of barge rate and water level.

Grain sales in the states along the Mississippi waterways at day $t$, $sales_t$, is used as an IV for $Q_t$ as barge quantities may be endogenous. Grain sales should correlate with barge movements on the rivers since markets along the rivers depend on the Mississippi transportation system for export, but sales is not directly correlated with barge rates. The correlations between sales and barge rates are 0.19 and 0.36 between barge quantities and barge rates. Monthly sales data are calculated as a product of the percent marketed of corn, soybeans and wheat by month and annual volume produced in states\footnote{The states include Illinois, Arkansas, Indiana, Iowa, Kentucky, Louisiana, Minnesota, Mississippi, Missouri, Ohio, Tennessee, and Wisconsin.} that rely on the river transport system. All data used to get the monthly sales data are obtained from Economic Research Service (ERS) of USDA. Monthly sales data are transformed by dividing the sales by the number of days in each month into daily data. Figure 3.6 shows the time series of transformed daily barges unloaded in New Orleans Region and grain sales from October 2005 to March 2013.
Figure 3.6: Number of Barges Unloaded in NOLA and Grain Sales for 2005 - 2013
3.5 Methodologies and Results

To investigate the effect of water level changes on barge rates for shipping grain from St. Louis to NOLA, regressions of barge rates on water levels are performed. Scatter plots of barge rates and water levels at St. Louis in Figures 3.7 and 3.8 show a piecewise relationship between the two can be considered. They show negative relationship between MISL and barge rates when water level is less than about stage 15 feet and the benchmark is about 30 ft for OHCA. No obvious relationships can be observed from the figures when they are above the benchmarks. A preliminary model is considered as:

\[
\ln(r_t) = \alpha + \beta_0 w_{sl}^t + \phi_0 d_{st}^t + \delta_0 d_{st}^t w_{sl}^t + \text{seasonality} + \epsilon_t, \quad (3.2)
\]

where \( d_{st}^t = 0 \) for \( w_{st}^t < w_{sl}^0 \) and \( d_{st}^t = 1 \) for \( w_{st}^t > w_{sl}^0 \). Benchmark water level at St. Louis, \( w_{sl}^0 \), is the turning point where the relationship between barge rates and water levels changes. At the benchmark point \( w_{sl}^0 \), barge rates in the two segments should be equal and there should be a discontinuity. Therefore, with restrictions that barge rates should be equal when \( d_{st}^t = 1 \) and \( d_{st}^t = 0 \) at benchmark water level \( w_{sl}^0 \), model 3.2 reduces to:

\[
\ln(r_t) = \alpha + \beta_0 w_{sl}^t + \delta_0 d_{st}^t (w_{st}^t - w_{sl}^0) + \text{seasonality} + \epsilon_t. \quad (3.3)
\]

Model 3.3 is estimated without including the quantity of barges at \( t \), \( Q_t \), which is a proxy
for the quantity of grain shipped from St. Louis to New Orleans. A model is estimated via 2SLS with $Q_t$ included:

$$
\ln(r_t) = \alpha + \beta_{sl}^{st} w_{sl}^{st} + \delta_{sl}^{st} (w_{sl}^{st} - w_0^{st}) + \gamma_0 \ln(Q_t) + \text{seasonality} + \epsilon_t, 
$$

(3.4)

where the logarithm of total grain sales $\ln(sales_t)$ is an IV for $\ln(Q_t)$. Table 3.3 shows the estimation results for both models. Model 3.3 is estimated first with $w_0^{sl} = 12$ selected by minimizing SSE with only integers for $w_0^{sl}$ considered. Model 3.4 is then estimated via 2SLS with the same benchmark water level of 12 ft. Note that the estimates of water level effects of $\beta_{sl}^{st}$ and $\gamma_0^{st}$ are very close in the two models. Although t-tests on the significance of each term may not be accurate due to possible serial correlation in residuals, the estimates still are consistent. Therefore, subsequent estimations of various models drop the quantity term, only with water levels at MISL and OHCA and their dynamic effects and seasonal components considered.

Seasonal components in both models are Fourier series. It is composed of three pairs of $\sin$ and $\cos$ terms, where $\text{seasonality} = \theta'_i \psi_{i,t} = \sum_{s=1}^{3} (\lambda_s \cos \frac{2\pi s t}{365} + \phi_s \sin \frac{2\pi s t}{365})$. For ease of notation, $\text{seasonality}$ is used in models to represent the seasonal components. The number of pairs is selected by backward selection starting with five pairs of components. The F-test used in backward selection of the seasonal components is calculated after serial correlation is removed by quasi-differencing. Three pairs of $\sin$ and $\cos$ terms.
Table 3.3: Estimates of Model 3.3 and 3.4

<table>
<thead>
<tr>
<th>Model</th>
<th>( \ln(r_t) = \alpha + \beta_0 w_{sl}^t + \delta_0 d_t^s (w_{sl}^t - w_{0}^s) + \text{seasonality} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficients</td>
<td>( \alpha )</td>
</tr>
<tr>
<td>Estimates</td>
<td>2.6756***</td>
</tr>
<tr>
<td></td>
<td>(0.0135)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>( \ln(r_t) = \alpha + \beta_0 w_{sl}^t + \delta_0 d_t^s (w_{sl}^t - w_{0}^s) + \gamma_0 \ln(Q_t) + \text{seasonality} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficients</td>
<td>( \alpha )</td>
</tr>
<tr>
<td>Estimates (OLS)</td>
<td>2.8127***</td>
</tr>
<tr>
<td></td>
<td>(0.1319)</td>
</tr>
<tr>
<td>Estimates (IV)</td>
<td>3.1789***</td>
</tr>
<tr>
<td></td>
<td>(0.5158)</td>
</tr>
</tbody>
</table>

\( w_{0}^s = 12 \) (unit: river stage). * Significant at the 10% level. ** Significant at the 5% level. *** Significant at the 1% level. \( \text{Adj} - R^2 = 0.4975 \) for model 3.3 and 0.4977 for model 3.4.
are included at the 5% significance level. Specifically, quasi-differencing is applied as indicated below. To simplify the notation, vector $Y_t$ represents the dependent variable and matrix $X_t$ represents the independent variables at time $t$. The variable $\epsilon_t$ is the residual from OLS regression in equation 3.5.

$$Y_t = \Theta X_t + \epsilon_t$$  \hspace{1cm} (3.5)$$

$$\epsilon_t = \rho \epsilon_{t-1} + u_t,$$  \hspace{1cm} (3.6)

where $u_t$ is serial uncorrelated. The quasi-difference is taken by

$$Y_t - \rho Y_{t-1} = \Theta (X_t - \rho X_{t-1}) + (\epsilon_t - \rho \epsilon_{t-1})$$  \hspace{1cm} (3.7)

and $\rho$ is replaced by its estimated regression coefficient in equation 3.6. Models estimated below are differenced if serial correlation is detected.

As argued in previous sections, water from the Upper Mississippi and Ohio River contribute flow to the river south of St. Louis. Both contemporaneous water levels at MISL and OHCA are included in the model as shown in equation (3.8):

$$\ln(r_t) = \alpha + \beta_0 w_{sl} + \delta_{sl} d_t (w_{sl} - w_{0,sl}) + \beta_{oh} w_{oh} + \delta_{oh} d_t (w_{oh} - w_{0,oh}) + \text{seasonality} + \epsilon_t,$$  \hspace{1cm} (3.8)

where $d_t = 0$ for $w_t < w_{0,i}$ and $d_t = 1$ for $w_t > w_{0,i}$, where $i = \{sl, oh\}$. Benchmark points in the relationships between barge rates and water levels are represented by $w_{0,i}$ for each river $i$.

Benchmark points $w_{0,i}$ are searched by minimizing SSE in an OLS regression before quasi-differencing. Serial correlation in residuals is detected since the coefficient $\hat{\rho}$, the first-order autocorrelation of residuals is high. Quasi-differences of variables are taken to reduce serial correlation such that conventional t and F tests can be performed to test the significance of explanatory variables. Results are shown in table 3.4.

SSE is minimized when the thresholds for water level stages are 11 ft at MISL and 30 ft at OHCA. Although t-tests on each term in significant at the 5% level except the intercept and $\delta_{0,oh}$, joint tests show that the contemporaneous water levels at MISL and OHCA are jointly significant at the 1% level. Insignificance of the water level at MISL and at OHCA may be attributed to high variance since water levels at MISL and OHCA are highly correlated, with a correlation of 0.70.
Water level effects on barge rates can be dynamic. Constant low water levels can affect barge rates more significantly and the length of information available in the market can have different effects on prices. In the following models, dynamic effects of water levels are considered as shown in the following equation:

\[
\ln(r_t) = \alpha + \sum_{j=0}^{p} [\beta_{sl}^j w_{t-j} + \delta_{sl}^j d_{t-j}(w_{t-j} - w_0^sl) + \beta_{oh}^j w_{t-j} + \delta_{oh}^j d_{t-j}(w_{t-j} - w_0^oh)] + \text{seasonality} + \epsilon_t, \tag{3.9}
\]

where \(p\) is total number of lagged water levels and \(j\) denotes the j-th lagged water level included in the model. It is assumed that the benchmark water levels \(w_i^0\) for \(i = \{sl, oh\}\) are the same for all time periods \(t - j\) and \(j = \{0, 1, 2\}\). The number of lags in the model is selected by forward selection method. Joint tests suggest that the number of lags, \(p\), is two at the 5% level.

Table 3.5 and 3.6 show the estimation results for models with once-lagged and twice-lagged water levels included. It shows that the contemporaneous water level effect at MISL is insignificant and the once-lagged water level effect is significant at the 5% level. The contemporaneous and the once-lagged effects at MISL are jointly significant at the 1% level. At OHCA, the contemporaneous water level is significant at the 10% level, and the once-lagged effect is not. Both of contemporaneous and once-lagged water levels are jointly significant at the 1% level. The tests indicate that both contemporaneous and once-lagged water levels at both locations should be included in the model.

The estimation result for model 3.9 with \(p = 2\) is shown in Table 3.6. Although only twice-lagged water level effect at MISL is significant at the 5% level, joint tests show that contemporaneous, once-lagged and twice-lagged water level effects at both locations are significant at the 5% level. Therefore, all terms should be included in the model.

The Table 3.7 shows the estimates of “long-run” water level effects below and above benchmarks at MISL and OHCA. When the water level at MISL is below stage 12 ft, the “long-run” effect of water level at MISL is the sum of \(\beta_j^{sl}\) for \(j = \{0, 1, 2\}\), which is -0.0151, and when the MISL water level is above 12 ft, the “long-run” effect is the sum of \(\beta_j^{sl}\) and \(\delta_j^{sl}\) for \(j = \{0, 1, 2\}\), which equals 0.0023. At OHCA, the “long-run” effect of water level is 0.0018 when the water level is below the benchmark of 38 ft and 0.0136 when the water level is above 38 ft. The standard errors of the “long-run” effects calculated by delta method show that water level below the benchmark of 12 ft at MISL
and water level above the benchmark of 38 ft at OHCA are significant at the 5% level. It means that when the MISL water level is below the benchmark of 12 ft, if the water level decreases by 1 foot and it remains for next two days, the barge rate at St. Louis will increase by 1.51%. If the barge rate is 14 dollars per ton, then it means an increase of 0.2114 dollars per ton. For a fully-loaded barge of corn, which is 1,500 tons, it is an increase of about 317 dollars per barge. The increase is even more remarkable for a tow that pushes about 25 barges. It will be an increase of 7,928 dollars of shipping cost. If the average corn price at St. Louis is 7 dollars per bushel, then the increase is 0.085% of the value of corn transported. At OHCA, a one-foot increase in the water level will cause 1.36% increase of barge rate at St. Louis, which is also 286 dollars per barge of increase for a fully-loaded barge. It is a 7,140 dollars increase for a tow that pushes 25 barges and about 0.076% of value of the corn transported. The water level at OHCA is 25 ft on average and also the flood stage is 40 ft, so when the water level is above 38 ft, it is possibly the flood that disrupt barge transportation and subsequently causes barge rates to increase.
Table 3.4: Model with only contemporaneous water levels

\[ \ln(r_t) = \alpha + \beta_0^{sl} w_t^{sl} + \delta_0^{sl} d_t^{sl}(w_t^{sl} - w_0^{sl}) + \beta_0^{oh} w_t^{oh} + \delta_0^{oh} d_t^{oh}(w_t^{oh} - w_0^{oh}) + \text{seasonality} + \epsilon_t \]

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>( \alpha )</th>
<th>( \beta_0^{sl} )</th>
<th>( \delta_0^{sl} )</th>
<th>( \beta_0^{oh} )</th>
<th>( \delta_0^{oh} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimates</td>
<td>2.6446***</td>
<td>-0.0054</td>
<td>0.0033</td>
<td>-0.0009</td>
<td>0.0084***</td>
</tr>
<tr>
<td></td>
<td>(0.0469)</td>
<td>(0.0028)</td>
<td>(0.0035)</td>
<td>(0.0017)</td>
<td>(0.0030)</td>
</tr>
</tbody>
</table>

\( H_0 : \beta_0^{sl} = \delta_0^{sl} = 0 \quad F = 2.8978 \quad Pval = 0.0556 \)
\( H_0 : \beta_0^{oh} = \delta_0^{oh} = 0 \quad F = 5.9692 \quad Pval = 0.0027 \)
\( H_0 : \delta_0^{sl} = \delta_0^{oh} = 0 \quad F = 4.6529 \quad Pval = 0.0097 \)
\( H_0 : \beta_0^{sl} = \delta_0^{sl} = \beta_0^{oh} = \delta_0^{oh} = 0 \quad F = 4.1725 \quad Pval = 0.0023 \)

\( w_0^{sl} = 11 \) and \( w_0^{oh} = 30 \) (unit: ft). * Significant at 10% significance level. ** Significant at 5% significance level. *** Significant at 1% significance level. \( \hat{\rho} = 0.9242 \) for quasi-difference.
Table 3.5: Model with up to 1 lagged water levels

\[ \ln(r_t) = \alpha + \beta_{0l} w_{tl} + \delta_{0l} d_{t}(w_{tl} - w_{0l}) + \beta_{0h} w_{th} + \delta_{0h} d_{t}(w_{th} - w_{0h}) \\
\beta_{1l} w_{tl-1} + \delta_{1l} d_{t-1}(w_{tl-1} - w_{0l}) + \beta_{1h} w_{th} + \delta_{1h} d_{t-1}(w_{th} - w_{0h}) + \text{seasonality} \]

<table>
<thead>
<tr>
<th></th>
<th>(\alpha)</th>
<th>(\beta_{0l})</th>
<th>(\delta_{0l})</th>
<th>(\beta_{0h})</th>
<th>(\delta_{0h})</th>
<th>(\beta_{1l})</th>
<th>(\delta_{1l})</th>
<th>(\beta_{1h})</th>
<th>(\delta_{1h})</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.6617***</td>
<td>-0.0038</td>
<td>0.0020</td>
<td>0.0004</td>
<td>0.0061</td>
<td>-0.0076***</td>
<td>0.0084**</td>
<td>-0.0007</td>
<td>0.0054</td>
<td></td>
</tr>
<tr>
<td>(0.0505)</td>
<td>(0.0029)</td>
<td>(0.0039)</td>
<td>(0.0020)</td>
<td>(0.0044)</td>
<td>(0.0028)</td>
<td>(0.0036)</td>
<td>(0.0020)</td>
<td>(0.0044)</td>
<td></td>
</tr>
</tbody>
</table>

\(H_0: \beta_{0l} = \delta_{0l} = 0\) \(F = 1.2131\) \(Pval = 0.2977\)
\(H_0: \beta_{1l} = \delta_{1l} = 0\) \(F = 3.7441\) \(Pval = 0.0240\)
\(H_0: \beta_{0l} = \delta_{0l} = \beta_{1l} = \delta_{1l} = 0\) \(F = 3.4979\) \(Pval = 0.0076\)
\(H_0: \beta_{0h} = \delta_{0h} = 0\) \(F = 2.9177\) \(Pval = 0.0545\)
\(H_0: \beta_{1h} = \delta_{1h} = 0\) \(F = 1.4781\) \(Pval = 0.2286\)
\(H_0: \beta_{0h} = \delta_{0h} = \beta_{1h} = \delta_{1h} = 0\) \(F = 4.3383\) \(Pval = 0.0018\)

\(w_{0l} = 11\) and \(w_{0h} = 32\) (unit: ft). * Significant at 10% significance level. ** Significant at 5% significance level. *** Significant at 1% significance level. \(\hat{\rho} = 0.9155\) for quasi-difference.
Table 3.6: Model with up to 2 lagged water levels

\[
\ln(r_t) = \alpha + \sum_{i=1}^{\{sl,oh\}} \left[ \beta_0^{sl} w_t^i + \delta_0^{sl}(w_t^i - w_0^i) + \beta_1^{sl}(w_{t-1}^i - w_0^i) + \beta_0^{oh} d_{i-1}^t + \delta_0^{oh} d_{i-1}^t (w_{t-1}^i - w_0^i) + \beta_2^{sl} w_{t-2}^i + \delta_2^{sl} d_{t-2}^i (w_{t-2}^i - w_0^i) \right]
\]

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$\beta_0^{sl}$</th>
<th>$\delta_0^{sl}$</th>
<th>$\beta_0^{oh}$</th>
<th>$\delta_0^{oh}$</th>
<th>$\beta_1^{sl}$</th>
<th>$\delta_1^{sl}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.6380***</td>
<td>-0.0042</td>
<td>0.0030</td>
<td>0.0011</td>
<td>0.0051</td>
<td>-0.0057*</td>
<td>0.0039</td>
</tr>
<tr>
<td>(0.0502)</td>
<td>(0.0027)</td>
<td>(0.0038)</td>
<td>(0.0017)</td>
<td>(0.0076)</td>
<td>(0.0028)</td>
<td>(0.0038)</td>
</tr>
<tr>
<td>$\beta_1^{oh}$</td>
<td>$\delta_1^{oh}$</td>
<td>$\beta_2^{sl}$</td>
<td>$\delta_2^{sl}$</td>
<td>$\beta_2^{oh}$</td>
<td>$\delta_2^{oh}$</td>
<td></td>
</tr>
<tr>
<td>-0.0003</td>
<td>0.0064</td>
<td>-0.0052*</td>
<td>0.0105***</td>
<td>0.0009</td>
<td>0.0003</td>
<td></td>
</tr>
<tr>
<td>(0.0018)</td>
<td>(0.0075)</td>
<td>(0.0027)</td>
<td>(0.0035)</td>
<td>(0.0013)</td>
<td>(0.0009)</td>
<td></td>
</tr>
</tbody>
</table>

$H_0$:
\[
\beta_0^{sl} = \delta_0^{sl} = 0 \quad F = 1.3475 \quad Pval = 0.2604
\]
\[
\beta_1^{sl} = \delta_1^{sl} = 0 \quad F = 2.1547 \quad Pval = 0.1165
\]
\[
\beta_2^{sl} = \delta_2^{sl} = 0 \quad F = 4.5243 \quad Pval = 0.0111
\]
\[
\beta_0^{oh} = \delta_0^{oh} = \beta_1^{oh} = \delta_1^{oh} = 0 \quad F = 2.7871 \quad Pval = 0.0255
\]
\[
\beta_1^{oh} = \delta_1^{oh} = \beta_2^{oh} = \delta_2^{oh} = 0 \quad F = 4.1695 \quad Pval = 0.0024
\]
\[
\beta_0^{sl} = \delta_0^{sl} = \beta_2^{sl} = \delta_2^{sl} = 0 \quad F = 3.0417 \quad Pval = 0.0166
\]
\[
\beta_0^{sl} = \delta_0^{sl} = \beta_1^{sl} = \delta_1^{sl} = \beta_2^{sl} = \delta_2^{sl} = 0 \quad F = 4.1275 \quad Pval = 0.0004
\]
\[
\beta_0^{oh} = \delta_0^{oh} = 0 \quad F = 1.1024 \quad Pval = 0.3325
\]
\[
\beta_1^{oh} = \delta_1^{oh} = 0 \quad F = 0.6487 \quad Pval = 0.5229
\]
\[
\beta_2^{oh} = \delta_2^{oh} = 0 \quad F = 0.5175 \quad Pval = 0.5962
\]
\[
\beta_0^{oh} = \delta_0^{oh} = \beta_1^{oh} = \delta_1^{oh} = \beta_2^{oh} = \delta_2^{oh} = 0 \quad F = 2.0481 \quad Pval = 0.0857
\]
\[
\beta_0^{oh} = \delta_0^{oh} = \beta_1^{oh} = \delta_1^{oh} = \beta_2^{oh} = \delta_2^{oh} = 0 \quad F = 2.2244 \quad Pval = 0.0388
\]

$w_0^{sl} = 12$ and $w_0^{oh} = 38$ (unit: ft). * Significant at 10% significance level. ** Significant at 5% significance level. *** Significant at 1% significance level. $\hat{\rho} = 0.9131$ for quasi-difference.
<table>
<thead>
<tr>
<th>Water Level</th>
<th>MISL</th>
<th>OHCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benchmark</td>
<td>&lt; 12 ft</td>
<td>&gt; 12 ft</td>
</tr>
<tr>
<td>Estimates</td>
<td>( \beta_0^{sl} + \beta_1^{sl} + \beta_2^{sl} )</td>
<td>( \beta_0^{sl} + \delta_0^{sl} + \beta_1^{sl} + \delta_1^{sl} + \beta_2^{sl} + \delta_2^{sl} )</td>
</tr>
<tr>
<td></td>
<td>(-0.0151^{***})</td>
<td>0.0023</td>
</tr>
<tr>
<td></td>
<td>(0.0035)</td>
<td>(0.0023)</td>
</tr>
</tbody>
</table>

* Significant at 10% significance level. ** Significant at 5% significance level. *** Significant at 1% significance level. Discrepancies are due to rounding errors.
Seasonal components are statistically significant at the 1% level even though it explains a small range of variations in barge rates as shown in Figure 3.9. As expected, the seasonal components plot shows that barge rates are high in late August and September and low in late March and April.

![Graph showing seasonality](image)

Figure 3.9: Seasonality in Model with up to 2 lags of water level

### 3.6 Conclusion

This paper analyzes the effects of water levels at MISL and OHCA on barge transport costs for shipping grains from St. Louis to New Orleans, LA. It captures drought and
flood effects on barge rates. Because water levels are exogenous and quantity shipped $Q$ does not affect estimates of water level effects, a regression of barge rates on water levels at MISL and OHCA and its derivative are analyzed. With up to two lags of water levels included in the model based on a forward selection method, joint F-tests suggest that the contemporaneous, once-lagged and twice-lagged water levels at both locations are significant at the 5% level. When the water level is below the benchmark water gage of 12 ft at MISL, the “long-run” effect of a one-foot decrease in water level is a 1.51% of increase in barge rates, which is equivalent to an average increase of 317 dollars in transport cost per fully-loaded corn barge. For a towboat that typically tows 25 barges on the lower Mississippi, the effect is even more remarkable, which is 7,928 dollars of increase in barge rates. The increase is about 0.085% of the value of corn transported if the corn price is at 7 dollars per bushel. At OHCA, when the water level is above 38 ft, it is possibly the flood that causes an increase in barge rates. The “long-run” estimate suggests that an one-foot increase of the water level at OHCA will cause 1.36% of increase in barge rates, which can be translated to 286 dollars per barge increase of barge rates for shipping corn to NOLA from St. Louis and about 0.076% of the value of corn transported.

Future work on the topic can consider a model with two benchmarks with regard to water levels since both low water levels and flood can disrupt barge transport and consequently increase shipping and operation costs. Additionally, instead of searching for the joint points of linear relationships that minimize the SSE first and removing serial correlation given the joint points, a search for joint points and quasi-differencing multiplier $\rho$ that minimize SSE can be performed simultaneously. Although the former can generate consistent results, the latter will have a better chance of achieving global efficiency.
References


