Spatial Price Integration in Competitive Markets with Capacitated Transportation Networks

by

(Ian) Yihang Zhu

A thesis submitted in conformity with the requirements
for the degree of Master of Applied Science
Department of Mechanical and Industrial Engineering
University of Toronto

Abstract

Spatial Price Integration in Competitive Markets with Capacitated Transportation Networks

(Ian) Yihang Zhu
Master of Applied Science
Department of Mechanical and Industrial Engineering
University of Toronto
2018

In this thesis, we examine the relationship between the equilibrium prices of spatially separated market participants and the transportation network that connects them; we focus in particular on isolating the effect that the transportation network can have on price integration. We first find that certain network structures and link costs can guarantee a bound on price differences between spatially separated participants when there are no capacity constraints. We extend this analysis to the case when there are binding transportation constraints, and generalize the effects of the transportation network on price integration using a time series decomposition. We then develop an empirical methodology, termed the submarket detection method (SDM), which can be used to infer when binding transportation constraints exist by analyzing regional pricing data from a competitive market, and conclude by using the SDM in a case study of the US gasoline market.

Dedicated to my parents.

Acknowledgements

I have been fortunate to have had the opportunity to work with many wonderful faculty members during my time as a Master's student. I would first like to thank my supervisor Timothy Chan for being an outstanding mentor, for believing in me, for pushing me, and for always looking out for me. You have taught me so much both academically and personally, and I feel very lucky that I can continue working with you in the coming years. I would also like to thank Michael Pavlin for being an incredible co-supervisor. Your constant encouragement, optimism, and patience has helped me overcome so many of the obstacles that I have faced over the last two years. Furthermore, I would like to thank Merve Bodur for her enthusiasm in exploring interesting research ideas with me, and for the career advice that she has given me. Finally, I would like to extend my thanks to John Birge, and my committee members Chi-Guhn Lee and Merve Bodur again for giving me valuable feedback on my work.

I would like to thank my labmates - Justin, Chris, Aaron, Rafid, Minha, Ben P., Ben L., Jonathan, Clara, Bing, Nasrin, Neal, Islay and Philip - for making the lab such a fun and stimulating environment to be in. Because of you, I am excited to come into the lab each day to work, to exchange ideas, or even just to hang out. I would also like to extend my thanks to all my friends outside the lab that have helped make Toronto feel like home in such a short amount of time.

I could not have done this without the endless support from Melissa. Thank you for always making me laugh, for taking care of me, and for putting up with me (especially when I'm grumpy). I am so grateful to have you in my life.

Finally, I would like to thank my parents for all that they have done for me. Thank you for working so hard to provide me with the opportunities that you never had, and for showing me that it is possible to achieve any goal with dedication and perseverance - I would not be where I am today without the lessons that you have taught me.

Contents

Chapter 1

Introduction

The gasoline market in the United States is currently one of the most competitive and integrated commodity markets in the world. On the supply side, extensive government deregulation has stimulated competition by allowing easy entry into the market for producers, retailers, and merchants. On the demand side, the transparency of price information provided by websites like Gasbuddy.com has enabled consumers to easily compare gas prices in local areas. As a result, individual retailers generally have little price-setting power, and it would be unusual to expect any one retailer to exhibit a significantly different price than its competitors. Furthermore, even if anomalies do exist at an individual level, we would expect citywide prices, which average over hundreds of retailers, to be similar. When considered over a time horizon, these factors would suggest that prices should be *integrated*, a general term to describe when prices move together. Studies of long-run gasoline prices have indeed shown that regional prices in the US are strongly integrated [26, 39].

The gasoline market, like many other energy markets, relies on a set of highly specialized transportation infrastructure to facilitate trade and competition between different market participants; in the gasoline market, this infrastructure takes the form of pipelines, tankers and trucks. However, the increasing demand for fuel in recent years combined with the lack of transportation capacity investment, often due to regulatory hurdles, have created a market that is frequently susceptible to binding transportation constraints. These transportation resources, when capacitated, have the ability to pull apart prices in regions where prices should theoretically be very similar [11, 37]. In particular, a transportation constraint can cause a subset of regions that use the resource to experience exaggerated price increases; we refer to this collection of regions as a *submarket*. The reliance of competitive prices and price integration on the transportation network can also be observed when there are disruptions in the network, such as those brought by a natural disaster, an accident or an infrastructural malfunction. In these cases, certain regional prices have been known to experience large unforseen surges [27, 30, 43]. A fully-functioning transportation network is so important for the market that significant resources have been put in place to secure and protect

the existing infrastructure, which have been cited as a potential target for terrorist attacks [36].

There is a clear connection between the underlying transportation network and price integration. Highly connected regions will support strong price integration since any arbitrage opportunities that may appear will be quickly arbitraged away by trading on these networks. Thus, the regional prices of any commodity market can be analyzed to infer exactly how connected different regions are. Many econometric methods, generally known as *co-integration tests*, do just this, by measuring how similarly prices at different regions behave, which can then be used to infer how connected different regions are. These tests are used ubiquitously in the gasoline market, as well as in many other commodity markets [26, 39, 46]. Like many econometric methods, the output of these tests are a set of summary statistics that can be used to describe price movements over a period. While these results may be sufficient to paint a good general picture of the market, they lack the ability to identify and characterize prices at specific time periods [5]. This results in the loss of a set of granular and potentially more useful information. For example, when analyzing prices generated from a market with capacity constraints, one might ask the following questions: can we determine when congestion occurs? how long do its effects last? which regions are impacted (i.e., forming a submarket) and by how much? Building a methodology that can answer these questions could yield valuable insights into the inference of underlying network constraints.

In this thesis, we attempt to build a methodology to identify when, which, and by how much different regions are affected by underlying bottleneck constraints. However, to do this effectively, we need to first have a clear understanding of *how* prices are distributed when there is (and when there isn't) congestion. In particular, we need to have a rigorous understanding of the relationship between the network structure, network parameters and the equilibrium prices.

In the first half of this thesis, we present a structural model of the network and examine the relationship between the transportation network and equilibrium prices. In particular, we provide a rigorous analysis of how the transportation network can directly influence how prices are distributed and the degree to which they are integrated. We present our findings in a time series context, in which prices are broken down into different components that each highlight the effects of network structure, transportation costs and transportation constraints. We use this time series representation to analyze exactly how regional prices are affected and distributed when there is (or isn't) congestion.

We then use the insights developed from the structural model to develop a methodology that will help us capture and characterize submarkets from the pricing data; we name the methodology the Submarket Detection Method (SDM). The main objectives of the SDM are to capture periods in which submarkets exist (i.e., periods in which large differences between regional prices appear), identify which, and to what degree different regions belong to the submarket. We then use the SDM to provide a case study of the US gasoline market using regional pricing data from 2016-2018, where we identify and estimate the effects of various bottleneck constraints on market prices.

1.1 Summary of Contributions

Our specific contributions are:

1. We provide a theoretical characterization of the effects of a capacitated network structure on price integration in a commodity market. We provide a novel bound on price integration that is a function of only network parameters, and also present a set of measures that can be used to evaluate the importance that different links have on price integration. We present a novel time series representation of equilibrium prices, in which each price is written as a function of different components capturing the effects of network structure, transportation costs and capacity constraints.

2. We develop a novel methodology that can accurately characterize price shocks given a set of time-series pricing data. Our method searches through different time periods and discovers when submarkets occur, the magnitude of the submarket, and the degree to which prices in different regions belong in the submarket.

3. We apply the SDM to historical gasoline pricing data in the United States to obtain the first large-scale analysis of congestion in the gasoline market. We use the SDM to evaluate the effect that several major pipeline disruptions had on regional prices. We also identify regions that may suffer from unobserved bottlenecks.

1.2 Outline of Thesis

The remainder of the thesis is organized as follows. In Chapter 2, we review the literature on models of competitive markets and the time series analysis methods that are used to analyze prices in competitive markets. In Chapter 3, we analyze the effect of network structure on prices in an uncapacitated setting. In Chapter 4, we extend our analysis to markets where there are capacity constraints. In Chapter 5, we present the Submarket Detection Method (SDM), which we use in Chapter 6 to analyze and estimate bottleneck effects in the southeastern US gasoline market. Finally, Chapter 7 concludes the thesis and discusses various future directions of this work.

Chapter 2

Literature Review

The two main fields in the literature to which our work relates are 1) structural models of competitive markets and 2) time series analysis methods.

2.1 Structural Models of Competitive Markets

For this thesis, the most relevant set of structural models of competitive markets are spatial price equilibrium (SPE) models. An SPE model is a model of a competitive market in which market participants are spatially separated and are connected by a transportation network [22, 41, 47, 48]. Mathematically, an SPE model is a mathematical programming model representing a perfectly competitive market, in which optimal market outcomes and equilibrium conditions can be derived directly from the optimality conditions (i.e., the Karush-Kuhn-Tucker conditions). The basic structure of an SPE model is one with an objective that maximizes consumer welfare minus transportation and production costs, and a set of production, consumption and flow balance constraints. The SPE model is a convex optimization model, where demand, supply, and transportation costs are all convex functions. A review of the different variants of spatial price equilibrium models is provided in [22]. These models are frequently used to predict market outcomes, i.e., equilibrium flows and prices, in many different commodity markets; within energy markets, SPE models have been used in the coal [21, 23], natural gas [18], crude oil [4], and petroleum industry [35]. More recently, SPE models have been used specifically to estimate the effects of capacity constraints on equilibrium flows and prices. For example, there are studies that estimate how the capacity constraints of the current natural gas pipeline network in the US and Europe, respectively, can affect prices given future estimates of demand [13, 33]. Finally, SPE models have also been used to study optimal investment decisions for transportation capacity [11, 42]. We differ from all these papers in that we focus on characterizing the theoretical relationship between the transportation network and price integration. In particular, previous research has focused on instance-specific outcomes,

for example, estimating equilibrium outcomes given a set of supply and demand functions. We differ by isolating the effect of the network on prices; for example, by analyzing how a network can influence price integration for arbitrary supply and demand functions.

2.2 Time Series Analysis Methods

The other major field that our work relates to is time series analysis methods. A large number of econometric methods have been proposed for commodity markets to measure the relationship between different geographically separated regions. The most relevant set of methods for this thesis are the ones that use regional price data to measure if, and to what degree, different regions are interacting and competing; these methods are known as co-integration tests. Co-integration tests measure whether or not price differences are stationary. When price differences between two regions are stationary, this by definition implies that any change in the price in one region is transmitted as an equal change in the price of the other region. Thus, co-integration has been used as a means to identify if different regions are interacting, either directly or indirectly. Many co-integration tests used in applications are for bivariate (pairwise) data [24, 25], but multivariate co-integration tests also exist [28, 40, 45] and measure co-integration between a set of three or more variables. Different variants of these tests exist, for example ones that account for threshold effects [3, 15, 32], i.e., co-integration is only measured once price differences exceed a certain threshold. These tests have been applied numerous times in the gasoline [26, 39] and natural gas [6, 12] markets. Lack of co-integration has also been proposed as a direct indication of bottleneck constraints in the underlying network [34].

Co-integration tests provide a set of summary statistics to describe relationships between regions. However, they cannot identify specific periods or time windows when the relationships are stronger or weaker [5]. In particular, while these methods may provide a rigorous means to measure price integration over a set of price data, they do not provide more detail into the relationships at specific periods for specific regions. We thus differ from these methods in our research objective to both identify specific time periods where price differences may arise, and to characterize regional relationships during these time periods.

For the detection and characterization of price shocks, we refer to the growing field of anomaly detection, which uses machine learning techniques to identify "unexpected behavior" in datasets. While anomaly detection techniques vary across a large range of different data types, significant contributions have been made in anomaly detection for time series data, both for univariate [9, 29, 31] and multivariate analysis [2, 8]. Anomaly detection in time series data largely revolves around identifying points or subsequences where a time series behaves unexpectedly. However, there are many different ways to define an anomaly and there is no definitive metric; a sequence in a time series may be considered an anomaly in one context but completely normal in another. Thus,

anomaly detection methods are almost always context specific and different for different application domains [7]. We contribute to this literature by developing a detection method where the anomaly being detected is a submarket. The insights from our theoretical model of a competitive market provide rigorous definitions and measurements for a submarket, which we use in the formulation of the SDM. To the best of our knowledge, this is the first detection method proposed to capture and estimate the effects of bottleneck constraints in the oil and gas markets.

Chapter 3

Network Structure and Prices

In this chapter, we study the relationship between the transportation network and equilibrium prices in the market. We begin by presenting a general model of a competitive market with transportation capacity constraints. We review how market outcomes can be derived using optimality conditions of the associated market allocation problem and establish fundamental equilibrium conditions between prices. We extend these conditions to show how the relationship between equilibrium prices in the market can be written as a direct function of the network structure and its parameters. In the absence of congestion, we show that only certain network structures can support and guarantee price integration; we call these networks *structurally integrated* networks. When the network is structurally integrated and uncongested, we show that price differences between any two nodes must be bounded, and the bound is defined by the transportation costs in the network. We also discuss how this bound can be used to evaluate the importance of certain links for price integration. The purpose of this chapter is to show that network structure directly influences the degree to which prices are integrated, and to examine the network conditions that can (or cannot) guarantee strong price integration.

3.1 Model and Preliminary Analysis

In this section, we develop a model of a competitive commodity market. Let the market be represented as a network with a set of nodes $\mathcal{N}$ and a set of directed links $\mathcal{E}$. Consumers are located at nodes $\mathcal{S}$ and producers are located at nodes $\mathcal{K}$, which together form a partition of $\mathcal{N}$. Each consumer node may be composed of many independent consumers in close spatial proximity (e.g., individual car owners purchasing gas, thereby defining a region, e.g., a city); the same is true for producer nodes. Each consumer node $s \in S$ obtains welfare $W_s(b_s)$ when consuming b_s units of the commodity, representing the aggregate welfare of individual consumers comprising node s. Similarly, each producer node $k \in K$ bears a production cost $W_k(b_k)$ for b_k units of the commodity

produced. We assume that the welfare function $W_s(\cdot)$ is strictly concave, increasing, and differentiable, while the cost function $W_k(\cdot)$ is convex, increasing and differentiable. The concavity and convexity assumptions are consistent with the standard diminishing marginal utility and diminishing return assumptions from the economics literature. For simplicity, we will frequently refer to consumer nodes and producer nodes simply as consumers and producers.

Nodes are connected by the set of transportation links $\mathcal{E}$. The variable f_{ij} represents the flow of the commodity from node i to j on link $(i,j) \in \mathcal{E}$. We use $I(i) = \{(n,i) \in \mathcal{E} \mid n \in N\}$ to denote the set of incoming nodes to i. Similarly, $O(i) = \{(i,n) \in \mathcal{E} \mid n \in N\}$ is the set of outgoing nodes from i. The flow on each link is non-negative, bounded above by the capacity of the link, u_{ij}, and has a non-negative, per-unit transportation cost of c_{ij}. We use $\mathcal{P}(i,j)$ to denote the set of paths from node i to j, and p_{ij}^q to denote the cost of a path $q \in \mathcal{P}(i,j)$, which is the sum of the costs on each link in q. For each pair of nodes (i,j) we let p_{ij}^* denote the cost of the minimum-cost path from i to j, i.e., $p_{ij}^* = \min\{p_{ij}^q \mid q \in \mathcal{P}(i,j)\}$. Finally, for a specific consumer $s \in \mathcal{S}$, we let the set $\mathcal{K}(s) \subseteq \mathcal{K}$ denote the set of producer nodes such that there exists a feasible flow path from $k \in \mathcal{K}$ to s. To ensure that the network is connected, we assume that $|\mathcal{K}(s)| \geq 1$ for all $s \in \mathcal{S}$.

Using the above notation, a competitive market can be modelled as the following welfare-maximizing market allocation problem:

$$
\begin{aligned}
\underset{\mathbf{f},\mathbf{b}}{\text{maximize}} \quad & \sum_{s \in \mathcal{S}} W_s(b_s) - \sum_{(i,j) \in \mathcal{E}} c_{ij} f_{ij} - \sum_{k \in \mathcal{K}} W_k(b_k) \\
\text{subject to} \quad & -b_s + \sum_{i \in I(s)} f_{si} - \sum_{j \in O(s)} f_{sj} = 0, \quad \forall s \in \mathcal{S}, \\
& b_k + \sum_{i \in I(k)} f_{ik} - \sum_{j \in O(k)} f_{kj} = 0, \quad \forall k \in \mathcal{K}, \\
& 0 \leq f_{ij} \leq u_{ij}, \quad \forall (i,j) \in \mathcal{E}, \\
& b_s \geq 0, \quad \forall s \in \mathcal{S}, \\
& b_k \geq 0, \quad \forall k \in \mathcal{K}.
\end{aligned}
\tag{3.1}
$$

The objective function maximizes welfare, which is the sum of consumer welfare minus transportation and production costs. The constraints include the standard flow balance equations, where consumers and producers withdraw and inject the commodity into the market and capacity constraints on flow. Given that $W_s(\cdot)$ and $W_k(\cdot)$ are strictly concave and convex functions, respectively, formulation (3.1) is a bounded, convex optimization problem. The optimality conditions of

problem (3.1), given below, provide us with the prices at equilibrium:

$$\lambda_s = W_s'(b_s) + \alpha_s, \quad \forall s \in \mathcal{S}, \tag{3.2a}$$
$$\lambda_k = W_s'(b_k) - \alpha_k, \quad \forall k \in \mathcal{K}, \tag{3.2b}$$
$$\lambda_j - \lambda_i = c_{ij} - \omega_{ij} + \nu_{ij}, \quad \forall (i,j) \in \mathcal{E}, \tag{3.2c}$$
$$\alpha_s, \alpha_k \geq 0, \quad \forall s \in S, \forall k \in \mathcal{K}, \tag{3.2d}$$
$$\omega_{ij}, \nu_{ij} \geq 0, \quad \forall (i,j) \in \mathcal{E}. \tag{3.2e}$$

The variables λ_s and λ_k are the dual variables corresponding to the two sets of flow balance constraints and represent the marginal cost of obtaining a unit of the commodity at the respective nodes; we will also refer to these variables as the *equilibrium prices* at the nodes. The variables α_s and α_k are dual variables of the lower bound constraints of b_s and b_k, respectively. The variables ω_{ij} and ν_{ij} are the dual variables corresponding to the lower and upper bound constraints on the flow variables, respectively. Since we are interested in studying the congestion effects of capacity constraints, whose marginal costs are represented by the variable ν_{ij}, we will refer to ν_{ij} as the *congestion surcharge*. We focus on Equation (3.2c), which establishes a connection between the prices at two nodes connected by a single link. If we sum this equation over a path q from node n_1 to n_2 that traverses edges in a set $\mathcal{E}_q$, then we get

$$\lambda_{n_2} - \lambda_{n_1} = \sum_{(i,j)\in\mathcal{E}_q} c_{ij} - \sum_{(i,j)\in\mathcal{E}_q} \omega_{ij} + \sum_{(i,j)\in\mathcal{E}_q} \nu_{ij} \tag{3.3}$$
$$= p^q_{n_1 n_2} - \sum_{(i,j)\in\mathcal{E}_q} \omega_{ij} + \sum_{(i,j)\in\mathcal{E}_q} \nu_{ij} \tag{3.4}$$

Since the variables $\omega_{ij}, \forall (i,j) \in \mathcal{E}$ are non-negative, Equation (3.4) can be reduced to,

$$\lambda_{n_2} - \lambda_{n_1} \leq p^q_{n_1 n_2} + \nu^q_{n_1 n_2} \tag{3.5}$$

where $\nu^q_{n_1 n_2} = \sum_{(i,j)\in\mathcal{E}_q} \nu_{ij}$ denotes the sum of the cumulative congestion surcharges of links on path q. Equation (3.5) is a fundamental no-arbitrage result of competitive markets. In particular, the equation states that the price at node n_2 cannot exceed the price at node n_1 plus the marginal cost of transporting a unit from n_1 to n_2. While this does establish a relationship between prices at different nodes in the network, it does not rule out the possibility that the price at a node is simply an arbitrary amount less than every other price in the network; Equation (3.5) fails to establish a strong connection between prices at any node in the network.

Nonetheless, for certain nodes, a stronger connection between prices can be established if nodes *participate* in the market. We define this term below:

Definition 3.1.1. A consumer $s \in \mathcal{S}$ participates in the market if $b_s > 0$ in the optimal market allocation.

When a consumer s participates in the market, a stronger connection can be established between the price at consumer s and another node in the network. This is expressed by the following lemma:

Lemma 3.1.1. *For every consumer $s \in \mathcal{S}$ that participates in the market, $\lambda_s = \min\{\lambda_k + p^q_{ks} + \nu^q_{ks} \mid k \in \mathcal{K}(s), q \in \mathcal{P}(k,s)\}$.*

Proof. We first prove that $\lambda_s \leq \min\{\lambda_k + p^q_{ks} + \nu^q_{ks} \mid \forall k \in \mathcal{K}(s),\ q \in \mathcal{P}(k,s)\},\ \forall s \in \mathcal{S}$. For any node s, there exists a path from every $k \in \mathcal{K}(s)$ to s, by the definition of $\mathcal{K}(s)$. Thus Equation (3.5) must hold for each of these producer-consumer pairs $(k,s),\ \forall k \in \mathcal{K}(s)$, i.e., $\lambda_s \leq \lambda_k + p^q_{ks} + \nu^q_{ks},\ \forall k \in \mathcal{K}(s)\ ,\ \forall q \in \mathcal{P}(k,s)$. This completes the first part of the proof. We now show that when $b_s > 0$ in the optimal allocation, $\lambda_s \geq \min\{\lambda_k + p^q_{ks} + \nu^q_{ks} \mid \forall k \in \mathcal{K}(s),\ q \in \mathcal{P}(k,s)\}\ \forall s \in \mathcal{S}$. We invoke equilibrium condition (3.4). Since there is positive consumption at the consumer node s, there must exist at least one path of positive flow from some producer $k^* \in K(s)$ to s in an optimal market outcome; we denote one of these paths as q^*. By complementary slackness, $\omega_{ij} = 0$ for all (i,j) in path q^*. From Equation (3.4), this implies that $\lambda_s - \lambda_{k^*} = p^{q^*}_{k^*s} + \nu^{q^*}_{k^*s}$. Rewriting this equation, we obtain $\lambda_s = \lambda_{k^*} + p^{q^*}_{k^*s} + \nu^{q^*}_{k^*s} \geq \min\{\lambda_k + p^q_{ks} + \nu^q_{ks} \mid q \in \mathcal{P}(k,s),\ k \in \mathcal{K}(s)\}$. □

Lemma 3.1.1 establishes a connection between the price of a consumer node and the price of a producer node. In particular, Lemma 3.1.1 states that the equilibrium price at a participating consumer node must be *equal* to the minimum marginal cost of production and transportation (which includes both inherent transportation cost and the congestion surcharge(s) on the path) over the set of producer nodes that the consumer is connected to. When there is no congestion in the network, Lemma 3.1.1 can be simplified to the following corollary:

Corollary 3.1.1. *For every consumer $s \in \mathcal{S}$ that participates in the market, $\lambda_s = \min\{\lambda_k + p^*_{ks} \mid k \in \mathcal{K}(s)\}$ when none of the links are congested.*

In this thesis, we are interested in characterizing the relationship between equilibrium prices and the underlying transportation network. However, since prices at nodes that do not participate in the market do not a have meaningful relationship with the prices in the rest of the network, we will make the following assumption for the rest of the thesis:

Assumption 3.1.1. *We assume that every consumer $s \in \mathcal{S}$ participates in the market.*

Another way to interpret this assumption is that it is an assumption on the structure of the welfare and cost functions of market participants. In particular, it assumes that the set of welfare

and cost functions ensures that each consumer participates in the optimal allocation; i.e., for any consumer in the market, the welfare gained from the first infinitesimally small unit consumed exceeds the cost of producing and transporting that unit. We will refer to the set of welfare and cost functions that guarantee consumer participation as *feasible* welfare and cost functions. We will now examine the relationship between the transportation network and equilibrium prices.

3.2 Structurally Integrated Markets

We begin by examining the connection between network structure and price integration. To best establish this connection, we will first examine price integration in the absence of transportation costs or capacity constraints. We begin by stating a condition on the transportation network structure.

Definition 3.2.1. A set of consumer nodes $\mathcal{S}_I \subseteq \mathcal{S}$ is *structurally integrated* if $\mathcal{K}(s) = \mathcal{K}(s')$, $\forall s, s' \in \mathcal{S}_I,\ s \neq s'$.

We say that a set of consumer nodes is structurally integrated if they share the same set of producers. This condition on the network structure is important because it allows us to establish a direct condition between the network structure and price integration in the absence of transportation frictions (i.e., costs and capacity constraints).

Lemma 3.2.1. *Assume that the transportation network has zero transportation costs and no capacity constraints. Then, the equilibrium price differences of any two consumers in a set of structurally integrated consumers is zero, for any set of feasible welfare and cost functions.*

Proof. Suppose that the set $\mathcal{S}_I$ represents the set of structurally integrated consumers. By Corollary 3.1.1, $\lambda_s = \min\{\lambda_k + p^*_{ks} | k \in \mathcal{K}(s)\},\ \forall s \in \mathcal{S}_I$. Since we are ignoring transportation costs, i.e., $p^*_{ks} = 0\ \forall k \in \mathcal{K}(s)\ \forall s \in \mathcal{S}_I$, then $\lambda_s = \min\{\lambda_k \mid k \in \mathcal{K}(s)\}$. Finally, since $\mathcal{K}(s)$ is the same for all $s \in \mathcal{S}_I$, the equilibrium prices λ_s must all be equal, i.e., $\lambda_s = \lambda'_s\ \ \forall s, s' \in \mathcal{S}_I$.

□

In other words, when two consumer nodes are structurally integrated, their equilibrium prices will always be equal for any arbitrary set of feasible welfare and cost functions when there are no transportation frictions. The statement also holds for a set of three or more structurally integrated consumers, and we say that *the market is structurally integrated* when this property holds for the entire set of consumer nodes $\mathcal{S}$. A set of nodes is thus structurally integrated when the transportation network is not a barrier to price integration, assuming the participants behave competitively and there are no trade frictions. We show the difference between structurally integrated and non-structurally integrated markets in the example below.

Example 3.2.1. *In this example, we will show that when a set of consumers is not structurally integrated, it is possible that price differences can arise even in the absence of transportation costs or capacity constraints. On the other hand, when a set of consumers is structurally integrated, there cannot be any price differences between different consumers when there are no transportation frictions.*

Consider the network shown in Figure 3.1a. We assume that transportation costs are zero and there are no capacity constraints on the network. In this network, there exist instances where different producer cost functions can lead to different prices between the consumers. For example, suppose both consumers have the same welfare function $W_s(b) = b^{1/2}$*, while the producers have different linear cost functions:* $W_{k_1}(b) = b$ *and* $W_{k_2}(b) = 2b$*. The equilibrium prices under this set of welfare functions are* $\lambda_{s_1} = 1,\ \lambda_{s_2} = 2$*, since consumer* s_2 *can only satisfy its demand from producer* k_2*, i.e., the more expensive producer.*

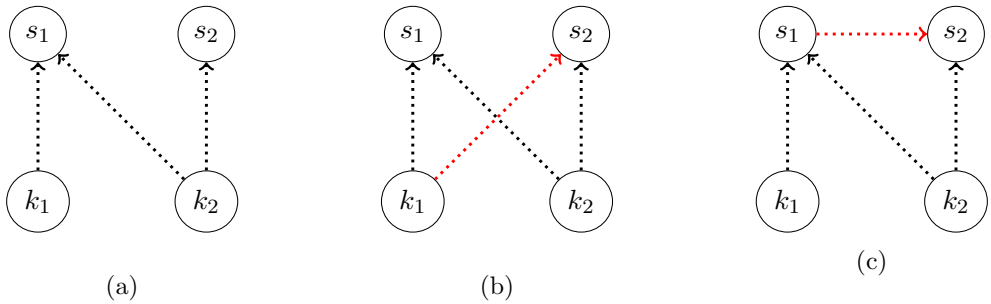

Figure 3.1: Examples of non-structurally integrated (a) and structurally integrated markets (b),(c).

However, when we add links (highlighted in red) that make the network structurally integrated, as shown in Figures 3.1b, 3.1c, consumer prices will be equal. In particular, for the markets in Figures 3.1b, 3.1c, the consumer prices will be $\lambda_{s_1} = \lambda_{s_2} = 1$ *since all consumers can buy from the cheaper producer (which is producer* k_1 *in this example).*

The definition of structural integration implies that if the set of consumers are connected to the same set of producers, then the equilibrium price of every consumer will be equal in the absence of transportation frictions. However, if this condition is not met, prices between consumers do not have to be equal, since some consumer may be able to buy from a cheaper producer that other consumers do not have access to.

3.3 Neutral Bands in Networks

In the previous section, we established a necessary and sufficient condition on the network structure to guarantee equal prices in the absence of transportation frictions. We now consider markets where transportation costs are non-zero but still in an uncapacitated setting. In this case, prices may

no longer be equal at each node. However, we now show that a structurally integrated market guarantees that the price differences will be bounded.

Theorem 3.3.1. *Two consumer nodes $s, r \in \mathcal{S}$ are structurally integrated if and only if the price difference between them are bounded according to the following equation for any set of feasible welfare and cost functions in the absence of capacity constraints:*

$$\min\{p^*_{ks} - p^*_{kr} \mid k \in \mathcal{K}(s)\} \leq \lambda_s - \lambda_r \leq \max\{p^*_{ks} - p^*_{kr} \mid k \in \mathcal{K}(s)\}. \tag{3.6}$$

Proof. ($\Rightarrow$) We first show that price differences between s and r cannot exceed this bound for any set of feasible welfare and cost functions. Since s and r are structurally integrated, by definition both s and r have access to the same set of suppliers $k \in \mathcal{K}(s)$. Since we assume that each consumer $s \in \mathcal{S}$ has positive consumption in the optimal market outcome, we let $\bar{k} \in \mathcal{K}(s)$ denote a producer such that there exists flow from $\bar{k}$ to s in the optimal market outcome. From Equation (3.4), this implies that $\lambda_s = \lambda_{\bar{k}} + p^*_{\bar{k}s}$. From Corollary 3.1.1, this also implies that $\lambda_r \leq \lambda_{\bar{k}} + p^*_{\bar{k}r}$. The equations can be combined into the following inequality: $\lambda_s - \lambda_r \geq p^*_{\bar{k}s} - p^*_{\bar{k}r}$. However, since we do not specify $\bar{k} \in \mathcal{K}(s)$, the following inequality must hold: $\lambda_s - \lambda_r \geq \min\{p^*_{ks} - p^*_{kr} | k \in \mathcal{K}(s)\}$. We can now make the same set of logical statements for the node r, thus bounding this inequality from the reverse direction. In doing so, we obtain the inequality $\lambda_s - \lambda_r \leq \max\{p^*_{ks} - p^*_{kr} | k \in \mathcal{K}(s)\}$, which completes the proof.

($\Leftarrow$) We argue by contrapositive; suppose the network is not structurally integrated. We show now that for any non-structurally integrated network, it is possible to find feasible welfare and cost functions such that there does not exist a finite bound between price differences. Without loss of generality, suppose producer $z \in \mathcal{K}(s)$ but $z \notin \mathcal{K}(r)$. Let the production cost functions be linear, of the form $W_k(b_k) = y_k b_k$, $\forall k \in \mathcal{K}(r)$ and $W_k(b_z) = y_z b_z$ for producer z. When the set of cost function coefficients satisfy $y_z + p^*_{zs} \leq \min\{y_k + p^*_{kr} \mid \forall k \in \mathcal{K}(r)\} + \epsilon$, the prices satisfy $\lambda_s \leq \lambda_r + \epsilon$ in the equilibrium. This is obtained by using Corollary 3.1.1, which states that $\lambda_s \leq y_z + p^*_{zs}$ and $\lambda_k = \min\{y_k + p^*_{kr} \mid \forall k \in \mathcal{K}(r)\}$. Thus, by considering cost functions that satisfy increasing values of ϵ, we can increasingly drive apart the values λ_s and λ_r; the price difference thus cannot be bounded. □

When two nodes are not structurally integrated, Theorem 3.3.1 implies that there are welfare functions that result in the price difference between these two nodes being arbitrarily large. However, when two nodes are structurally integrated, the price difference is bounded and, more importantly, the bound is characterized entirely by the network topology and link costs. This result does not imply that the equilibrium prices themselves are bounded; they can vary depending on the specific welfare functions. It is instead the *price difference* between nodes that is bounded, with the worst-case price difference being characterized by the bounds in Equation (3.6).

Lemma 3.3.1. *When the nodes $s, r \in \mathcal{S}$ are structurally integrated, the bound from Equation (3.6) is the tightest bound on the price difference for all feasible welfare and cost functions.*

Proof. To show that the bound from Equation (3.6) is the tightest, we show that for any $k \in \mathcal{K}(s)$, it is possible to come up with a set of cost functions such that k services both s and r in an optimal market outcome; in this outcome, $\lambda_s - \lambda_r = p^*_{ks} - p^*_{kr}$. Let k^* denote an arbitrary producer in the set $\mathcal{K}(s)$. Let $y_k,\ k \in \mathcal{K}$ be a set of linear cost functions coefficients that satisfy the following two conditions: 1) $y_{k^*} + p^*_{k^*s} \leq \min\{y_{k_o} + p^*_{k_o s} \mid \forall k_o \in \mathcal{K}(s) \backslash k^*\}$ and 2) $y_{k^*} + p^*_{k^*r} \leq \min\{y_{k_o} + p^*_{k_o r} \mid \forall k_o \in \mathcal{K}(s) \backslash k^*\}$. Since the cost functions are linear, $\lambda_k = y_k,\ \forall k \in \mathcal{K}$, and by Corollary 3.1.1, $\lambda_s = \min\{\lambda_k + p^*_{ks} | k \in \mathcal{K}(s)\} = \lambda_{k^*} + p^*_{k^*}$ and similarly, $\lambda_r = \min\{\lambda_k + p^*_{kr} | k \in \mathcal{K}(s)\} = \lambda_{k^*} + p^*_{k^*}$. Thus, for any arbitrary $k^* \in \mathcal{K}(s)$, it is possible to construct a set of cost functions such that the equation $\lambda_s - \lambda_r = p^*_{k^*s} - p^*_{k^*r}$ holds in equilibrium. Hence, the bound in Equation (3.6) must be the tightest bound for arbitrary cost functions (i.e., without making any assumptions on cost functions). □

The idea that there exists a bound on price differences is not new, and is referred to as a *neutral band* [20, 38]. A neutral band is defined as a region within which price differences can fluctuate without any arbitrage opportunities existing. This phenomenon has been examined in many different applications. For example, in finance, when there are positive transaction costs between securities, price differences can exist without creating any profit opportunity [16, 17]. In commodity markets, transportation and trade cost between two regions have similarly been shown to create regions within which prices differences can fluctuate [14, 19]. In our context, this is given by the following equation,

$$-c_{sr} \leq \lambda_s - \lambda_r \leq c_{rs} \tag{3.7}$$

when there exists a link from s to r, and a link from r to s; this is obtained from Equation (3.2c). In other words, the price differences between any two nodes must fall within the cost of transportation between the two nodes in any competitive market.

To the best of our knowledge, however, our result is the first to generalize the concept of a neutral band over an entire network (as opposed to a single link). In doing so, we obtain Theorem 3.3.1, which is a more general result that extends the neutral band idea to a set of nodes that are connected, without requiring direct connections between each pair of nodes. In particular, Equation (3.6) does not require that there exists any direct transportation methods between a pair of nodes for their prices to be linked to each other. A tight neutral band can exist between nodes even when they are not able to directly trade with one another; the prices at two consumer nodes are still related because they have the same set of shared producers and thus compete "indirectly". Furthermore, even if two nodes $r, s \in \mathcal{S}$ are directly connected, the bound in Theorem 3.3.1 can be strictly tighter than the bound from Equation (3.7). This is shown in the following example:

Example 3.3.1. *In this example, we show how Equation (3.6) can provide a strictly tighter bound between two consumer prices even when they are directly connected to each other. Suppose the market is represented by the network in Figure 3.2, where* $s1, s2$ *denote the two consumers in the market and* $k1, k2$ *denote the producers.*

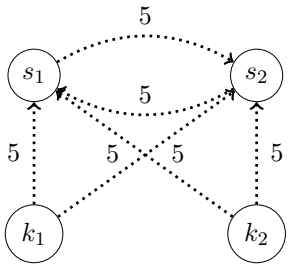

Figure 3.2: An example of a network where consumers are directly connected to each other.

Since there are direct links between s_1 *and* s_2 *of cost* 5 *units, Equation (3.7) suggests that the price difference between* s_1 *and* s_2 *cannot exceed 5 units. However, if we were to look at the network as a whole, i.e., consider the consumers' connectivity to the producers, Equation (3.6) would find that the prices must actually be equal in all instances, i.e.,* $\lambda_{s_1} = \lambda_{s_2}$*. Thus, this example highlights the importance of considering the entire network even when analyzing subsets of market participants.*

Theorem 3.3.1 provides a novel way of thinking about price integration between prices not just from how they may be directly connected, but rather by how they are positioned in the underlying network. We now interpret what it means for a pair of prices to have a large band width. For a large neutral band width to exist between s_1 and s_2, two conditions must be met: 1) there must be at least one producer $k_1 \in \mathcal{K}$ such that transportation from k_1 to s_1 is cheaper than transportation from k_1 to s_2, **and** 2) there exists a *different* $k_2 \in \mathcal{K} \backslash \{k_1\}$ such that transportation from k_2 to s_2 is cheaper than from k_2 to s_1. In particular, a large neutral band width can only exist when each consumer node has its own set of suppliers for which it is cheaper to trade with. We demonstrate this idea in the following example:

Example 3.3.2. *In this example, we show how the distribution of transportation costs on the network directly determines the degree to which prices are integrated.*

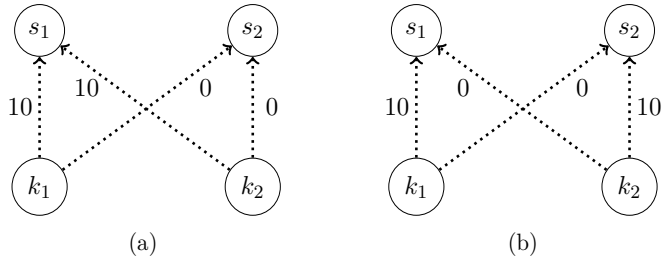

Figure 3.3: Two markets with the same network structure but different distribution of link costs.

In Figure 3.3a, even though s_2 can access all the suppliers for a cheaper value, this does not imply the existence of a large neutral band. In fact, in this particular example, the neutral band has width zero, and $\lambda_{s_1} = \lambda_{s_2} + 10$, i.e., the price at s_1 will always be exactly 10 units higher than at s_2. On the other hand, the width of the neutral band in Figure 3.3b is large; any $\lambda_{s_1}, \lambda_{s_2}$ such that $\lambda_{s_1} - \lambda_{s_2} \in [-10, 10]$ is feasible. This phenomenon occurs only because each consumer s_1 and s_2 has a producer (i.e., k_2 and k_1 respectively) for which it is cheaper to trade with.

The above example highlights the fact that it is not necessarily the magnitude of transportation costs that determine price integration between consumers in the network, but rather it is *how* the transportation costs are distributed across the network.

Equation (3.6) can also be used to assess the importance of a specific link on price integration between pairs of prices. In particular, we can compare the neutral band calculated between prices from the original network with one calculated from the network where a link is removed. Doing so will provide a worst-case estimate for how prices can be distributed when a certain link is disrupted or congested. Similarly, if a neutral band does not exist when the link is removed, then this is an immediate indication that the link is important for keeping the prices in the market integrated. We explore this idea in more detail in the following chapter.

In this chapter, we have shown that there is a clear relationship between network structure and price integration. We presented a simple necessary and sufficient condition on the network structure to *guarantee* that there exists some relationship between the prices in the market; this condition is referred to as structural integration. When a market is structurally integrated, we showed that it is the distribution of transportation costs on the network that determines how strongly prices are integrated, as measured by the neutral band defined in Equation (3.6). When the market has small neutral bands, the transportation network is sufficient to support strong price integration, and large demand and supply changes alone cannot cause major shifts in price differences. In the next chapter, we explore how price integration can be substantially weakened when there are transportation constraints in the network.

Chapter 4

Capacity Constraints and Prices

In this chapter, we analyze the relationship between equilibrium prices when there are binding capacity constraints in the underlying market. Unlike the previous chapter, we choose to perform this analysis specifically in the context of a time series model. In our time series model, each price is broken down into separate components describing the effect of the network structure, network parameters and transportation constraints in the market. By analyzing prices in this context, we will be able to directly identify the components that are affected by congestion and predict how prices will behave when different links are congested. These insights will be directly used in the methodological and empirical sections, where we attempt to infer the effects of congestion on different regions using only time series data of consumer prices.

We assume that the market is structurally integrated in this chapter. We will first define a set of values that will be used frequently in this chapter.

Definition 4.0.1. For any pair of nodes $r, s \in \mathcal{S}$, let ρ_{rs} denote the midpoint of the neutral band defined in Equation (3.6), i.e.,

$$\rho_{rs} = (\min\{p^*_{ks} - p^*_{kr} | k \in \mathcal{K}\} + \max\{p^*_{ks} - p^*_{kr} | k \in \mathcal{K}\})/2. \tag{4.1}$$

Similarly, let the value h_{rs} denote half the width of the band, i.e.,

$$h_{rs} = (\max\{p^*_{ks} - p^*_{kr} | k \in \mathcal{K}\} - \min\{p^*_{ks} - p^*_{kr} | k \in \mathcal{K}\})/2. \tag{4.2}$$

Thus, equation (3.6) can be rewritten as $\lambda_s - \lambda_r \in [\rho_{rs} - h_{rs} \ , \ \rho_{rs} + h_{rs}]$.

4.1 Uncongested Networks

We begin by representing the prices generated from an uncongested network as a time series; this is an extension of equation (3.6) to multiple periods.

Lemma 4.1.1. *Let the set $\lambda_s^t,\ s \in \mathcal{S}$ represent the set of consumer equilibrium prices at time t over a time horizon $\mathcal{T}$, generated from a market with arbitrary supply and demand functions but with a fixed transportation network. When none of the links are congested over the time period $\mathcal{T}$, the set of consumer prices can be described by the following time series:*

$$\lambda_s^t = \lambda_o^t + \rho_{os} + \epsilon_s^t \quad \forall s \in \mathcal{S},\ \forall t \in \mathcal{T}, \tag{4.3}$$

where $\epsilon_s^t \in [-h_{os}, h_{os}]$ and $o \in \mathcal{S}$ represents an arbitrary (root) node.

Proof. The result can be obtained directly from Equation (3.6). □

Equation (4.3) implies that prices in a structurally integrated and uncongested market can be perfectly expressed by an underlying trend λ_o^t, a time-invariant term ρ_{os} and a bounded noise term ϵ_s^t. Lemma 4.1.1 can also be interpreted as follows. Given any set of prices λ_s^t, all variation among the set of prices can be fully captured by the value ρ_{os} and variables ϵ_s^t, where ϵ_s^t is bounded within the fixed region $[-h_{os}, h_{os}]$ defined strictly by the parameters of the network. Furthermore, as we had proved in Theorem 3.3.1, the band that ϵ_s^t is confined to be in is the *tightest* band width that can fully capture all the price variation generated from different feasible welfare and cost functions. We will now present a similar time series representation for prices generated from a congested network.

4.2 Congested networks

In this section, we present a functional form representation of prices generated from a network with a congested link. We first introduce notation and definitions that will be used in this section.

Recall that $\mathcal{P}(k, s)$ defines the set of directed paths from a node $k \in \mathcal{K}$ to a node $s \in \mathcal{S}$. We partition $\mathcal{P}(k, s)$ into the set $\mathcal{P}_e(k, s)$ and $\mathcal{P}_{-e}(k, s)$, which define the set of paths which include link e and the set of paths which do not include link e, respectively. Let p_{ks}^e and p_{ks}^{-e} denote the cost of the min-cost path from k to s in the set $\mathcal{P}_e(k, s)$ and $\mathcal{P}_{-e}(k, s)$, respectively. If the set $\mathcal{P}_e(k, s)$ is empty, let $p_{ks}^e = \infty$. Similarly, if $\mathcal{P}_{-e}(k, s)$ is empty, let $p_{ks}^{-e} = \infty$. Since we assume the network is structurally integrated, at most one of these two sets can be empty. We will now use these terms to examine and define the effect that a capacitated link can have on prices in the market.

Definition 4.2.1. A pair $(k, s), k \in \mathcal{K}, s \in \mathcal{S}$ *relies* on link e if $p_{ks}^e < p_{ks}^{-e}$. Similarly, if there exists at least one $k \in \mathcal{K}$ for a node $s \in \mathcal{S}$ where this is true, we say that consumer s relies on e.

Note the subtlety in the definition; a pair (k, s) does not necessarily rely on e if e is on a min-cost path from (k, s). The pair (k, s) only relies on e if it is on a min-cost path between (k, s), *and* there does not exist another min-cost path between (k, s) that does not use e. When (k, s) relies on link e, congestion of this link will directly affect the cost of trade between the producer k and

consumer s. Nonetheless, when the congestion surcharge on link e reaches a certain threshold in which it becomes cheaper for a consumer s to obtain from k using a path in $\mathcal{P}_{-e}(k,s)$ (i.e., a path that does not use link e), further increases in the congestion surcharge will no longer affect the transportation cost between k and s. This is defined rigorously below as the *replacement cost* of link e for a producer-consumer pair (k,s).

Definition 4.2.2. The *replacement cost* of a link e on a producer-consumer pair (k,s) is the minimum amount of congestion surcharge that can be added to link e such that the pair (k,s) no longer relies on e. The replacement cost is denoted by $RC_e(k,s)$ and is defined as:

$$RC_e(k,s) = \max\{p_{ks}^{-e} - p_{ks}^{e}, 0\} \tag{4.4}$$

When $p_{ks}^{-e} - p_{ks}^{e} < 0$, the min-cost path between k and s does not include link e (by definition), and thus the replacement cost $RC_e(k,s)$ is equal to zero. Otherwise, the congestion surcharge on link e can increase the price between (k,s) by at most the additional cost of using an alternative path that does not include link e. If such an alternative path does not exist, i.e., $P_{-e}(k,s) = \emptyset$, then the $RC_e(k,s) = \infty$. We now extend the notion of replacement cost to the entire set of producers $k \in \mathcal{K}$.

Definition 4.2.3. Let $RL_e(s), RU_e(s)$, and $RW_e(s)$ denote the minimum replacement cost, maximum replacement cost, and replacement window of a link e on a node s, respectively. These are defined by the following equations:

$$RL_e(s) = \min\{RC_e(k,s) \mid k \in \mathcal{K}\} \tag{4.5a}$$

$$RU_e(s) = \max\{RC_e(k,s) \mid k \in \mathcal{K}\} \tag{4.5b}$$

$$RW_e(s) = \begin{cases} RU_e(s) - RL_e(s) & \text{if } \ RL_e(s) < \infty \\ 0 & \text{if } \ RL_e(s) = \infty \end{cases} \tag{4.5c}$$

Finally, we use these definitions to rigorously define a *submarket*.

Definition 4.2.4. Let $\mathcal{S}_e \subseteq \mathcal{S}$ denote the set of consumer nodes that rely on link e. The set $\mathcal{S}_e$ forms a *submarket* when link e is congested and $\nu_e > 0$. We refer to ν_e as the *value* of the submarket.

In words, a submarket is a collection of nodes which experience a positive congestion surcharge as a result of a congested link. We now use these definitions to present the main result of this chapter: a time series representation of equilibrium prices generated from a capacitated network.

Theorem 4.2.1. *Let λ_s^t denote a set of equilibrium prices that is generated from a network over the time period $\mathcal{T}$, in which a link e is congested and has congestion surcharge ν_e^t, $t \in \mathcal{T}$. Let o denote a node that does not rely on link e. The set of consumer equilibrium prices can be fully captured in this equation:*

$$\lambda_s^t = \lambda_o^t + \rho_{os} + \epsilon_s^t + \min\{\nu_e^t, RL_e(s)\} + r_s^t, \quad \forall s \in \mathcal{S},\ \forall t \in \mathcal{T} \tag{4.6}$$

where $\epsilon_s^t \in [-h_{os}, h_{os}]$, $r_s^t \in [0, RW_e(s)]$.

Proof. We fix the root node o to be a node that does not rely on e; without loss of generality, we set this node as the source node of link e. For each node $s \in \mathcal{S}$, we break the case down to when $\nu_e^t \leq RL_e(s)$, and when $\nu_e^t > RL_e(s)$.

Case I: We first examine the case when $\nu_e^t \leq RL_e(s)$ for some $t \in \mathcal{T}$. We claim that when this is true, all variation can be captured by the following equation:

$$\lambda_s^t = \lambda_o^t + \rho_{os} + \epsilon_s^t + \nu_e^t \quad \text{where } \epsilon_s^t \in [-h_{os}, h_{os}]. \tag{4.7}$$

By definition of $RL_e(s)$, the cost of the shortest path between any producer $k \in \mathcal{K}$ and s must be equal to $p_{ks}^* + \nu_e^t$ (i.e., ν_e^t is not high enough to replaced by any cheaper min-cost paths). In any optimal market outcome, there must be some positive flow from a producer $k \in \mathcal{K}$ to consumer s; we denote this producer as k^*. Equilibrium Equation (3.4) implies that $\lambda_s^t = \lambda_{k^*}^t + p_{k^*s}^* + \nu_e^t$. Similarly, Corollary 3.1.1 states that $\lambda_o^t \leq \lambda_{k^*}^t + p_{k^*o}^*$. On the other hand, since λ_o^t must also obtain some positive flow from one of the producers, there exists $k^o \in \mathcal{K}$ such that $\lambda_o = \lambda_{k^o} + p_{k^o o}^*$ and $\lambda_s \leq \lambda_{k^o} + p_{k^o s}^* + \nu_e^t$. The following equation must hold must hold: $\min\{(p_{ks}^* + \nu_e^t) - p_{ko}^* | k \in \mathcal{K}\} \leq \lambda_s - \lambda_o \leq \max\{(p_{ks}^* + \nu_e^t) - p_{ko}^* | k \in \mathcal{K}\}$, or written slightly differently, $\nu_e^t + \rho_{os} - h_{os} \leq \lambda_s - \lambda_o \leq \nu_e^t + \rho_{os} + h_{os}$. All possible price differences between λ_s^t and λ_o^t can be captured by Equation (4.7).

Case II: When $\nu_e^t > RL_e(s)$, we claim that the series can be represented as:

$$\lambda_s^t = \lambda_o^t + \rho_{os} + RL_e(s) + \epsilon_s^t + r_s^t \quad \text{where } \epsilon_s^t \in [-h_{os}, h_{os}], r_s^t \in [0, RW_e(s)]. \tag{4.8}$$

Thus, we need to show that all variation between prices can be fully captured by the bounded variables ϵ_s^t and r_s^t. To prove this statement, we look at the maximum price difference a congested link e can create between a pair of nodes (o, s). We determine this by analyzing the neutral band when the link is completely removed, which is equal to:

$$\min\{p_{ks}^{-e} - p_{ko}^* | k \in \mathcal{K}\} \leq \lambda_s - \lambda_o \leq \max\{p_{ks}^{-e} - p_{ko}^* | k \in \mathcal{K}\}. \tag{4.9}$$

Since this neutral band denotes the worst price differences that a congested link e can create for pairs (o, s), we need to show that the entire neutral band can be captured by the bounded variables ϵ_s^t and r_s^t. To do this, we show that $RL_e(s) + \rho_{os} - h_{os} \leq \min\{p_{ks}^{-e} - p_{kr}^{*} | k \in \mathcal{K}\}$ and $RL_e(s) + \rho_{os} + h_{os} + RW_e(s) \geq \max\{p_{ks}^{-e} - p_{ko}^{*} | k \in \mathcal{K}\}$. We begin by showing that the first inequality holds: $RL_e(s) + \rho_{os} - h_{os} = \min\{p_{ks}^{-e} - p_{ks}^{e} | k \in \mathcal{K}\} + \min\{p_{ks}^{*} - p_{ko}^{*} | k \in \mathcal{K}\} \leq \min\{p_{ks}^{-e} - p_{ks}^{e} + p_{ks}^{*} - p_{ko}^{*} | k \in \mathcal{K}\} = \min\{p_{ks}^{-e} - p_{ko}^{*} | k \in \mathcal{K}\}$.Now we show that the second inequality also holds: $RL_e(s) + \rho_{os} + h_{os} + RW_e(s) = RL_e(s) + \rho_{os} + (RU_e(s) - RL_e(s)) = \max\{p_{ks}^{*} - p_{ko}^{*} | k \in \mathcal{K}\} + RU_e(s) = \max\{p_{ks}^{*} - p_{ko}^{*} | k \in \mathcal{K}\} + \max\{p_{ks}^{-e} - p_{ks}^{*} | k \in \mathcal{K}\} \geq \max\{p_{ks}^{*} - p_{ko}^{*} + p_{ks}^{-e} - p_{ks}^{*} | k \in \mathcal{K}\} = \max\{p_{ks}^{-e} - p_{ko}^{*} | k \in \mathcal{K}\}$.

□

Equation (4.6) provides a functional form representation of prices in a congested network in which the influence of the network features on price integration can be easily observed. We use three examples of increasing complexity to discuss the implications of Theorem 4.2.1.

4.2.1 Example I: networks without alternative paths

In the first example, we examine networks that are directed spanning trees, in which there is a single unique path from every producer to every consumer; we refer to these networks as networks with no alternative paths. We use the Figure 4.1 as an example of such a network.

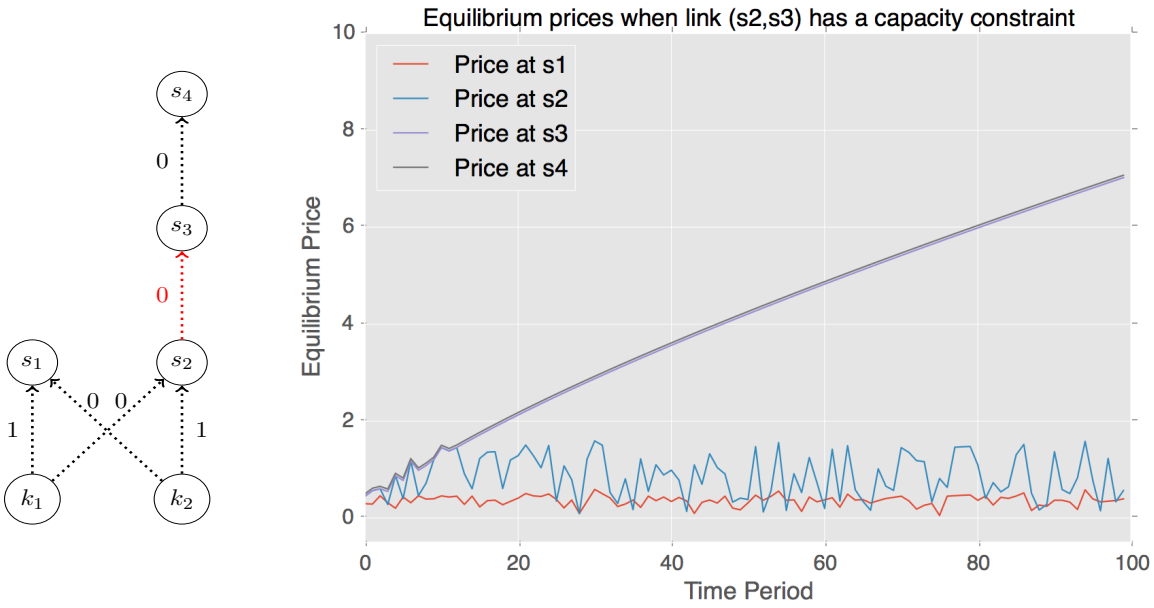

Figure 4.1: An example of prices generated from a congested network that has no alternative paths.

In this network, the link $e = (s_2, s_3)$ has a fixed capacity constraint while all other links are

uncapacitated, and the demand at node s_3 increases with time. At each instance $t \in [0, 100]$, new values of x and y are randomly generated between [0,0.2], and the producer cost functions are set as $W_k(b_{k_1}) = xb_{k_1}^2$ and $W_k(b_{k_2}) = yb_{k_2}^2$. The welfare functions remain as $W_s(b) = b^{1/2}$ for nodes s_1, s_2, s_4 for all $t \in [0, 100]$, whereas the welfare function is $W_s(b) = t^{3/4}b^{1/2}$ for node s_3.

We note that even though it is the increasing welfare function (i.e., increasing demand) of s_3 that is intensifying the congestion surcharge, the equilibrium price at s_4 is nevertheless affected since the node s_4 relies on link (s_2, s_3). We can also observe the prices of the nodes that do not rely on link (s_2, s_3) satisfying the neutral band of $\lambda_{s_1}^t - \lambda_{s_2}^t \in [-1, 1],\ \forall t \in \mathcal{T}$.

Using this example, the time series representation of equilibrium prices in a network with no alternative paths can be simplified to the following equation:

$$\lambda_s^t = \lambda_o^t + \rho_{os} + \epsilon_s^t + \nu_e^t \cdot I_{es}, \quad \forall s \in \mathcal{S}, \forall t \in \mathcal{T}$$

where node $o \in \mathcal{S}$ denotes a node that does not rely on link e, $\epsilon_s^t \in [-h_{os}, h_{os}]$, and $I_{es} = 1$ if s relies on e and $I_{es} = 0$ otherwise. The proof is simple: for the consumer nodes $s \in \mathcal{S}$, $RL_e(s) = \infty$ for nodes that rely on e, $R_e(s) = 0$ for nodes that do not, and $RW_e(s) = 0\ \forall s \in \mathcal{S}$. Thus, when the network has no alternative paths, a congested link will cause all nodes downstream of the link to suffer the same congestion value, which can be unbounded.

There are many network structures that might not have alternative links; for example, a perfect bipartite graph between producers and consumers. In this setting, congestion on any link will only affect the sink node of the link, since a perfect bipartite graph implies that each node uniquely and independently uses its own set of transportation links.

4.2.2 Example II: networks with alternative paths but low RWs

We now examine the distribution of prices when there are alternative links, but when the replacement window is small. We use Figure 4.2 as an example of such a network. The setup is the same as in Example 4.1: the welfare and cost functions are generated in the same manner, and the link (s_2, s_3) is congested. However, in this example there is an additional replacement link for node s_4. In this network, $RL_e(s_1) = RL_e(s_2) = 0$, $RL_e(s_4) = 3$, $RL_e(s_3) = \infty$, and $RW_e(s) = 0\ \forall s \in \mathcal{S}$.

In this example, we observe that as the demand at node s_3 increases, the congestion surcharge on link (s_2, s_3) increases as well. However, once the congestion surcharge on link (s_2, s_3) exceeds 3 units, the price at s_4 is no longer affected by an increasing ν_e^t; on the other hand, the congestion surcharge that can be incurred by s_3 is unbounded. Since the replacement window is zero, we see that when the value $\nu_e^t > 3$, the value of s_4 reaches a new equilibrium where it is now fixed to be 3 units higher than s_1 and s_2, but the price pair differences between s_1, s_2 and s_4 still fluctuates within the same neutral band width as before. In other words, when $\nu_e^t > 3$, a new, fixed price difference between the nodes s_1, s_2 and s_4 is reached, and the noise generated from instance-specific

welfare and cost functions can still bound by the original neutral band width.

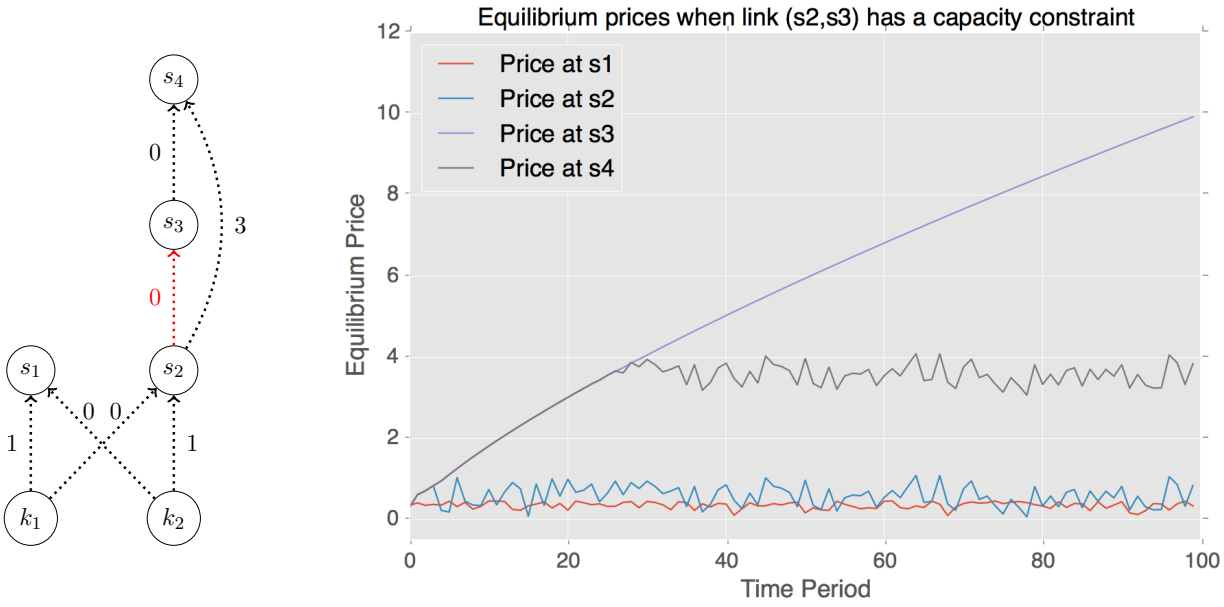

Figure 4.2: An example of prices generated from a network with an alternative path.

In this example, the set of equilibrium prices $\lambda_s^t,\ \forall s \in \mathcal{S}, t \in \mathcal{T}$ can thus be written as:

$$\lambda_s^t = \lambda_o^t + \rho_{os} + \epsilon_s^t + \min\{\nu_e^t, RL_e(s)\}, \quad \forall s \in \mathcal{S},\ \forall t \in \mathcal{T}$$

with $o \in \mathcal{S}$ once again denoting a node that does not rely on e (e.g. node s_1), and $\epsilon_s^t \in [-h_{os}, h_{os}]$.

4.2.3 Example III: networks with high RWs

We will now show that when a node has a large replacement window for a capacitated link, the price at the node can exhibit significant volatility when the link is congested. We will use the network shown in Figure 4.3 to demonstrate this phenomenon. In this example, the producer cost functions are $W_k(b_{k_1}) = xb_{k_1}^2$ and $W_k(b_{k_2}) = yb_{k_2}^2$, where x is randomly generated from [0,1] and y is randomly generated from [0,0.2] at every instance $t \in [0, 100]$. The welfare functions are $W_s(b) = b^{1/2}$ for consumer nodes s_1 and s_2, and $W_s(b) = t^{3/4}b^{1/2}$ for nodes s_3 and s_4. In this network, $RW_e(s_4) = 3$. This implies that the node s_4 can have an additional window of 3 units to vary within when link e is congested and when the congestion surcharge exceeds $RL_e(s_4) = 3$. We thus observe that as the congestion surcharge grows, significantly more volatility is observed in the price at s_4 than the prices at s_1 and s_2.

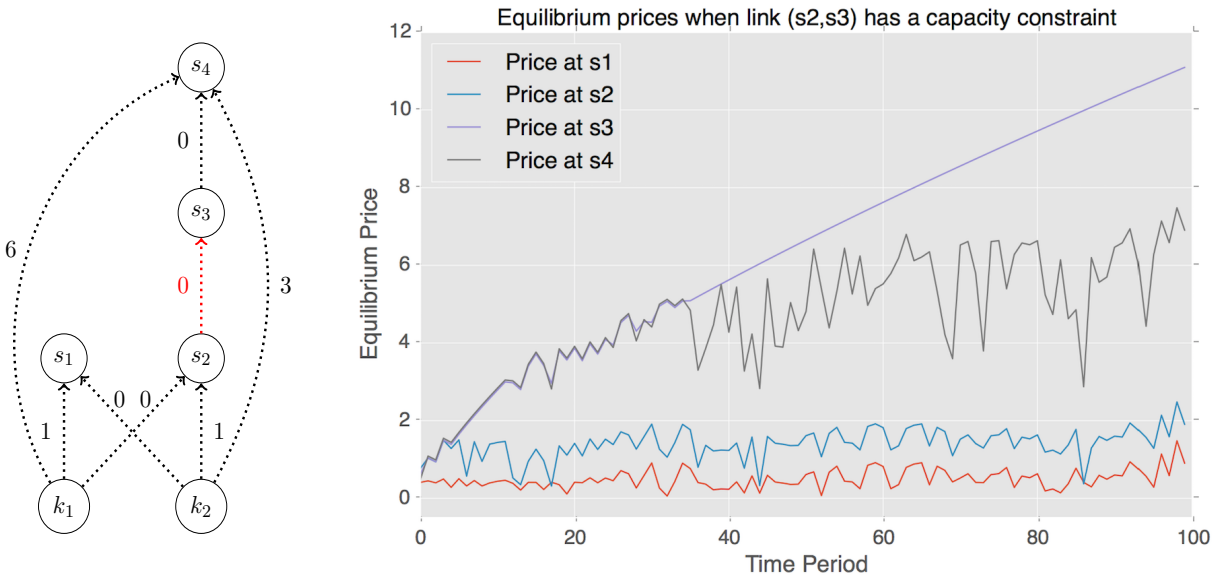

Figure 4.3: An example of prices generated from a network where nodes have large replacement windows for the congested link.

4.3 Summary of Findings

In summary, Chapter 3 and Chapter 4 highlight the relationship between the transportation network and equilibrium prices in any competitive market. When the network is not congested, price integration between market participants is directly related to the distribution of minimum-cost paths between producer-consumer pairs. When there is congestion in the network, the rest of the network will determine how nodal prices respond to the congestion. When there exists alternative paths that can replace the congested link, prices may not fully bear the full weight of the shadow price of the constraint. However, even when there are paths that can replace a congested link, major differences in the cost of replacement paths can generate increased volatility in prices.

Chapter 5

Capturing Submarkets

In this section, we discuss how to capture and characterize submarkets from a set of equilibrium prices generated from an fixed, capacitated network. As discussed in the previous section, a submarket can arise when there is a positive congestion surcharge on a congested link. Similarly, each node in the submarket can have a different replacement cost for the congested link, above which the congestion surcharge can no longer (fully) increase the prices. We use these insights to propose a methodology, termed the submarket detection method (SDM), which is built to capture these submarket characteristics, namely: 1) when a submarket appears, 2) the value of the submarket (i.e., the congestion surcharge of the unobserved transportation constraint) 3) which nodes are in the submarket, and 4) how the price at each node in the submarket responds to the submarket value. However, since the SDM is built to analyze energy (and other competitive commodity) markets, we will now use terminology that is more consistent with the application; in particular, we will refer to nodes as regions, and equilibrium prices as regional prices. We treat these terms to be synonymous.

5.1 Measures of Fit

Let $\lambda_s^t,\ s \in \mathcal{S}, t \in \mathcal{T}$ denote a set of observed consumer prices; in particular, λ_s^t is now no longer a variable, but instead a value in the observed dataset. We begin by defining a measure of variation among a set of prices; we denote this as the market integration value.

Definition 5.1.1. Given a set of commodity prices $\lambda_s^t,\ s \in \mathcal{S}, t \in \mathcal{T}$, the *market integration value* (V), is given by the following equation:

$$V = \min_{\epsilon, \rho_s, \eta^t} \{\ \epsilon\ \text{ s.t. }\ |\lambda_s^t - \eta^t - \rho_s| \leq \epsilon,\ \forall s \in S,\ \forall t \in \mathcal{T}\}. \tag{5.1}$$

The value of V is a measure of the maximum deviation between any two price series over the observed period of time and is related to the neutral band of the underlying network, as described

in the following lemma.

Lemma 5.1.1. *If the set of prices λ_s^t are generated from a market where none of the links are congested in $t \in \mathcal{T}$, then $V \leq \max\{h_{rs} \mid r, s \in \mathcal{S}\}$.*

Proof. The variables $\eta^t,\ \forall t \in \mathcal{T}$ and $\rho_s,\ s \in \mathcal{S}$ are free in Equation (5.1). Thus, if we can find a set of values for η^t and ρ_s such that $V \leq \max\{h_{rs} \mid r, s \in \mathcal{S}\}$, we are done. First, set $\eta^t = \lambda_o^t$ for an arbitrary region $o \in \mathcal{S}$. Let $\rho_s = \rho_{os}$. Thus, from Equation (4.3), $V \leq \max\{h_{os} \mid s \in \mathcal{S}\} \leq \max\{h_{rs} \mid r, s \in \mathcal{S}\}$. □

Theoretically, if we knew that the prices were generated from a network that was not congested over the observed time period, V would be less than or equal to half the width of the maximum neutral band between any two regions. On the other hand, if the set of prices was generated from a network that had congested links generating congestion surcharges that pull apart prices, the value of V could be increased significantly. In this setting, if we were able to capture and remove the congestion values from the set of prices, we would expect to be able to reduce the value V. We define the value of V after removing the effects of congestion to be the *optimized* market integration value.

Definition 5.1.2. The *optimized market integration value* (V^o) is defined as:

$$V^o = \min_{\epsilon, \rho_s, \eta^t} \{\ \epsilon \ \text{ s.t. } \ |\lambda_s^t - \eta^t - \rho_s - \hat{\mu}_s^t| \leq \epsilon,\ \forall s \in \mathcal{S}, \quad \forall t \in \mathcal{T}\} \tag{5.2}$$

where $\hat{\mu}_s^t$ denotes the estimated congestion surcharge affecting region $s \in \mathcal{S}$ in period $t \in \mathcal{T}$.

The *optimized* market integration value defines a measure of price integration when we remove the (estimated) effects of congestion. When a market is consistently plagued by congestion, the value V^o will likely be significantly lower than V. This provides the motivation behind the submarket detetection method, which provides a structured way of estimating the values $\hat{\mu}_s^t$ such that V^o is minimized.

5.2 Submarket Detection Method

We present the submarket detection method below. We first introduce the variables and parameters used for the SDM, then present the model formulation. The SDM takes a set of prices as input and attempts to discover and characterize submarkets. We briefly remind the reader that a submarket is a collection of regions in the market that collectively experience a price increase as a result of an underlying bottleneck in the transportation network; we define the value of a submarket to be the estimated congestion surcharge of the unobserved bottleneck. The first input paramater of the model is the number of submarkets that we want to consider; this is similar to the k in k-means clustering. Let R denote this input parameter, and let $\mathcal{R} = \{1, 2, ..., R\}$ denote the indices for each

submarket. For each submarket $r \in \mathcal{R}$, let the variable ν_r^t represent the value of submarket r at time $t \in \mathcal{T}$. For each region $s \in \mathcal{S}$ and each submarket $r \in \mathcal{R}$, let the variable δ_{sr} represent the replacement cost of submarket r for region s. The variable δ_{sr} thus determines whether or not a region belongs in a submarket, and if so, to what degree. Let variables μ_{sr}^t represent the congestion surcharge that region s receives from submarket r. When $\nu_r^t \leq \delta_{sr}$, i.e., the submarket value is below the replacement cost, and thus $\mu_{sr}^t = \nu_r^t$. On the other hand, when $\nu_r^t \geq \delta_{sr}$, the variable $\mu_{sr}^t = \delta_r^t$; region s will no longer be affected by increasing values of ν_r^t. Finally, let the variable μ_s^t denote the cumulative congestion surcharge of all the submarkets $r \in \mathcal{R}$ on s, i.e., $\mu_s^t = \sum_{r \in \mathcal{R}} \mu_{sr}^t$. The submarket detection method is given by the following formulation:

$$\underset{\eta^t, \rho_s, \epsilon, \nu_r^t, \pi_{sr}, \delta_{sr}}{\text{minimize}} \quad \epsilon \tag{5.3a}$$

$$\begin{aligned}
\text{subject to} \quad & \mu_s^t \leq \lambda_s^t - \eta^t - \rho_s + \epsilon, && \forall r \in \mathcal{R},\ \forall s \in \mathcal{S},\ t \in \mathcal{T}, && \text{(5.3b)}\\
& \mu_s^t \geq \lambda_s^t - \eta^t - \rho_s - \epsilon, && \forall r \in \mathcal{R},\ \forall s \in \mathcal{S},\ t \in \mathcal{T}, && \text{(5.3c)}\\
& \mu_s^t = \sum_{r \in \mathcal{R}} \mu_{sr}^t && \forall s \in \mathcal{S}, t \in \mathcal{T}, && \text{(5.3d)}\\
& \mu_{sr}^t \leq \nu_r^t, && \forall r \in \mathcal{R},\ \forall s \in \mathcal{S},\ t \in \mathcal{T}, && \text{(5.3e)}\\
& \mu_{sr}^t \leq \delta_{sr}, && \forall r \in \mathcal{R},\ \forall s \in \mathcal{S},\ t \in \mathcal{T}, && \text{(5.3f)}\\
& \mu_{sr}^t \geq \nu_r^t - (1 - \pi_{sr}^t) \cdot M, && \forall r \in \mathcal{R},\ \forall s \in \mathcal{S},\ t \in \mathcal{T}, && \text{(5.3g)}\\
& \mu_{sr}^t \geq \delta_{sr} - \pi_{sr}^t, && \forall r \in \mathcal{R},\ \forall s \in \mathcal{S},\ t \in \mathcal{T}, && \text{(5.3h)}\\
& \pi_{sr}^t = \{0, 1\}, && \forall r \in \mathcal{R},\ s \in \mathcal{S}, && \text{(5.3i)}\\
& \eta^t, \mu_{sr}^t, \nu_r^t \geq 0, && \forall r \in \mathcal{R},\ s \in \mathcal{S},\ t \in \mathcal{T}. && \text{(5.3j)}
\end{aligned}$$

Constraints (5.3b) and (5.3c) define whether the observed price at a region $s \in S$ can be captured by η^t, ρ_s^t and ϵ. If not, the remainder will be captured by the value μ_s^t, which represents the congestion surcharge that region s is experiencing at time t. The value μ_s^t is equal to $\sum_{r \in \mathcal{R}} \mu_{sr}^t$, the aggregate sum of all submarkets that affect region s, defined by constraint (5.3d).

The set of constraints (5.3e), (5.3f), (5.3g), (5.3h) determine the relationship between each submarket $r \in \mathcal{R}$ and each region $s \in \mathcal{S}$. The variable δ_{sr} sets the upper bound on how much submarket r can affect s. When the submarket value ν_r^t is below this threshold, $\mu_{sr}^t = \nu_r^t$. This is defined by the constraints (5.3e), (5.3g), with $\pi_{sr}^t = 1$. On the other hand, when ν_r^t exceeds the threshold δ_{sr}, $\mu_{sr}^t = \delta_{sr}$. This is defined by the constraints (5.3f), (5.3h), with $\pi_{sr}^t = 0$.

We set $M = \max\{\ \lambda_r^t - \lambda_s^t \ \mid r, s \in \mathcal{S},\ t \in \mathcal{T}\ \}$. Since M measures the biggest price difference between any two nodes across the entire observed period, ν_r^t would never exceed this value. Finally, the optimized objective value of the SDM is denoted by the value V^o.

5.2.1 Comparison with Time Series Representation

We now discuss the connection between the SDM model and the time series representation presented in Chapter 4. We recall Equation (4.6), i.e.,

$$\lambda_s^t = \lambda_o^t + \rho_{os} + \epsilon_s^t + r_s^t + \min\{\nu_e^t, RL_e(s)\}, \quad \forall s \in \mathcal{S},\ \forall t \in \mathcal{T}, \tag{5.4}$$

where $\epsilon_s^t \in [-h_{os}, h_{os}],\ r_s^t \in [0, RW_e(s)]$. The variables in this equation capture all possible price differences among any two pairs of nodes in the network for a specific congested link e. We design the SDM to solve a model that mimics this equation. To avoid confusion, we remind the reader that all symbols below, with the exception of R (number of submarkets) and λ_s^t (the observed regional prices), are free variables. The following model is a more compact way of expressing the SDM:

$$\begin{aligned}
\underset{\eta^t, \rho_s, \epsilon_s^t, \epsilon, \nu_r^t, \delta_{sr}}{\text{minimize}} \quad & \epsilon \\
\text{subject to} \quad & \lambda_s^t = \eta^t + \rho_s + \epsilon_s^t + \sum_{r \in \mathcal{R}} \min\{\nu_r^t, \delta_{sr}\}, && \forall s \in \mathcal{S},\ \forall t \in \mathcal{T}, \\
& \epsilon_s^t \in [-\epsilon, \epsilon], && \forall s \in \mathcal{S},\ \forall t \in \mathcal{T}, \\
& \nu_r^t, \delta_{sr} \geq 0, && \forall r \in \mathcal{R},\ \forall s \in \mathcal{S},\ \forall t \in \mathcal{T}.
\end{aligned} \tag{5.5}$$

The similarity between the constraints in the SDM and the time series representation of prices should be obvious. The SDM solves a model that mimics the Equation (4.6) to infer the values of ν_r^t (value of submarket r) and δ_{sr} (replacement cost of a region s for a submarket r), while minimizing the absolute error between prices. However, there are a few key differences. First, it is possible that more than one (unobservable) link gets congested in the underlying transportation network during the observed sample period, and the input parameter R specifies a "guess" for how many. Secondly, Equation (4.6) uses a node o (that does not rely on a congested link) as a root node. However, given a dataset of regional prices, we cannot pre-identify which region is or is not affected by congestion, and thus use instead a variable η^t to represent a root node (or an underlying trend).

5.3 Reducing Overfitting

In the SDM, it is easy to see that there exist a large number of variables compared to the size of the input dataset. It thus becomes easy for the model to overfit the data. Overfitting in this model is typically observed by two qualities in the model output: 1) a submarket is active (i.e., has positive value) for extended periods of time but at very low values, and 2) regions have very low but non-zero δ_{rs} values. These two factors can make the results less interpretable by obscuring when submarkets are actually active and which regions are in a submarket. We thus propose two additional sets of constraints to help reduce overfitting.

5.3.1 Constraining the Number of Active Submarket Periods

The first restriction we propose is a constraint on the number of periods for which a submarket can be active for. Mathematically, this is a constraint on the number of periods $t \in \mathcal{T}$ such that $\nu_r^t > 0$ for each submarket $r \in \mathcal{R}$. The set of constraints used to impose this restriction is presented below:

$$\nu_r^t \leq \omega_r^t \cdot M, \qquad \forall r \in \mathcal{R},\ t \in \mathcal{T}, \tag{5.6a}$$

$$\sum_{t \in \mathcal{T}} \omega_r^t \leq x_r \cdot T, \qquad \forall r \in \mathcal{R}, \tag{5.6b}$$

$$\omega_r^t \in \{0, 1\}, \qquad r \in \mathcal{R},\ t \in \mathcal{T}. \tag{5.6c}$$

The binary variable ω_r^t is used to indicate when a submarket can be active; if $\omega_r^t = 0$, then $\nu_r^t = 0$, and if $\omega_r^t = 1$, constraint (5.6a) becomes redundant, and ν_r^t is not constrained. The number of time periods that ω_r^t can be nonzero is constrained by constraint (5.6a), where the parameter x_r is used to indicate the maximum proportion of time periods that submarket r can be active for. The parameter M is the same as before, and T denotes the number of time periods in the dataset. Finally, instead of fixing x_r with an input parameter, we can set x_r as a variable and put it as a regularization term in the objective function.

5.3.2 Constraining the Number of Submarket Allocations

The second restriction that we propose is a constraint on the number of submarkets that each region can be allocated to. Since the variable δ_{sr} decides if (and to what degree) a region is part of a submarket, we impose a constraint on the number of submarkets for which δ_{sr} can exceed zero. This is given by the following set of constraints:

$$\delta_{sr} \leq \gamma_{sr} \cdot M, \qquad \forall r \in \mathcal{R},\ s \in \mathcal{S}, \tag{5.7a}$$

$$\sum_{r \in \mathcal{R}} \gamma_{sr} \leq y_s, \qquad \forall s \in \mathcal{S}, \tag{5.7b}$$

$$\gamma_{sr} \in \{0, 1\}, \qquad \forall r \in \mathcal{R},\ s \in \mathcal{S}. \tag{5.7c}$$

Once again, the parameter M is the same as before. When $\gamma_{sr} = 0$, the variable $\delta_{sr} = 0$. When $\gamma_{sr} = 1$, the constraint (5.7a) becomes redundant. The parameter y_s is used to impose the maximum number of submarkets that region s can belong to. Instead of fixing y_s with an input parameter, we can set y_s as a variable and put it as a regularization term in the objective function.

Chapter 6

An Application to the US Gasoline Market

In this chapter, we examine how the SDM can be used to analyze and estimate the effects of bottleneck constraints in the US gasoline market using retail gasoline prices. We first provide a brief introduction of the gasoline market and discuss specific insights that we can obtain using the SDM. We will then present a detailed case study of the southeastern market, depicted in Figure 6.1.

The transportation network underlying the US gasoline market is extremely vast and complicated. While pipelines are the most frequently used mode of transportation, a significant portion of gasoline is also transported by rail, trucks, and tankers, which are typically more flexible; in fact, rail is consistently cited as an alternative use of transport whenever pipelines become congested [44]. The vast number of transportation methods create a highly complex network in which capacity and flow data become near-impossible to obtain, and one where it becomes difficult to determine the dependence of regions on certain transportation resources. In such a setting, we will show how the SDM can be used to answer two important questions:

1. When there are well-documented cases of disruption, can we evaluate how large the congestion surcharge was, and how prices at different regions responded to the disruption?

2. When there are no well-documented cases of disruptions, can we still identify areas that may suffer from bottleneck constraints arising from day-to-day trading?

When there is a well-documented disruption, the SDM can be used to estimate the congestion surcharge that arose from the disruption. Similarly, the SDM also determines how prices at different regions responded to the price shock, which can elucidate how dependent different regions are on a specific transportation resource, and how connected different regions are by secondary modes of

transport, i.e., alternative pipelines, rail and trucks. On the other hand, when there are no well-documented sources of disruption, we can use the SDM to highlight areas that may consistently suffer from transportation bottlenecks.

6.1 Case study: The Southeastern Gasoline Market

In this case study, we focus on the gasoline prices in the southeastern states in the US. We consider 10 adjacent states, from Texas to Virgina, which cover some of the areas of highest activity in the US petroleum industry. Within this market, pretroleum products flow unidirectionally from the refineries in the Gulf Coast to the southeastern and eastern states of the US. Over 50% of the nation's refining capacity is located right along the coast of Texas and Louisiana, and the cities to the east are serviced almost exclusively by the the refineries in the Gulf Coast [27]. The refineries feed into two main pipelines, the Colonial Pipeline and the Plantation Pipeline, which ship the gasoline to any number of locations between the Gulf Coast and the eastern-most cities of the US. The fact that all the cities in the region are serviced by the same set of producers and the close proximity of producers suggest that strong price integration should be expected in the absence of capacity constraints. However, the transportation infrastructure in this region is frequently operating at full capacity and is also susceptible to transportation disruptions caused by various weather-related natural disasters that frequent the region such as hurricanes [27]. We believe that these features of the market make for an interesting case study and opportunity to apply the SDM.

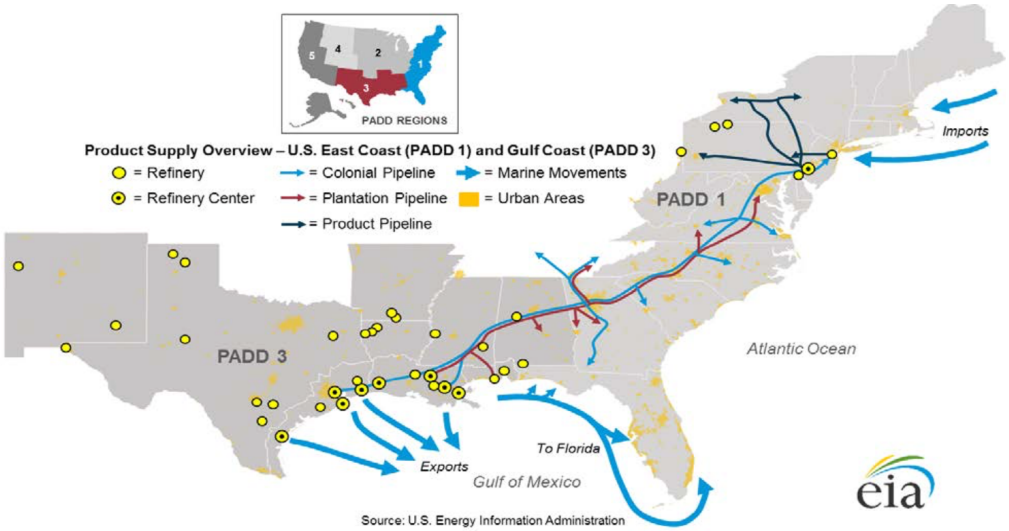

Figure 6.1: The flow of petroleum products in the southeastern US gasoline market. Source: [27]

6.1.1 Experimental Setup

We obtained a set of daily regional gasoline prices over the two-year period 2016-2018. The prices cover 20 cities in this region. The cities in the sample include the largest city for each state (according to US Census Bureau), and a few other densely populated cities in the region. A table of the cities along with basic summary statistics is provided in the Appendix. We first smooth the time series by using a centered moving-average with a window of three days. Due to the complexity of our model (i.e., one binary variable for each data point), we do not use the daily price but instead use prices from every three days. We split the dataset into half-year intervals: (01-01-2016 to 07-01-2016), (07-01-2016 to 12-31-2016), (01-01-2017 to 07-01-2017) and (07-01-2017 to 12-31-2017). These samples will be referred to as 2016H1, 2016H2, 2017H1 and 2017H2 for the rest of the chapter. For each sample, we set $R = 2$ (number of submarkets) and $x_r = 0.25, \forall r \in \{1, 2\}$ (maximum proportion of active periods of submarket r), and set a stopping criteria of two hours.

6.1.2 Empirical Results

We run the SDM for each half-year interval. Each submarket is plotted geographically below, along with the value of each submarket. We omit plotting the δ_{sr} values for ease of visualization; these are instead presented in the Appendix. We briefly remind the reader that the value of submarket r, i.e., ν_r^t for some time $t \in \mathcal{T}$, represents the congestion surcharge of an unobservable bottleneck, whereas the δ_{sr} values represent the maximum amount that the submarket value can add to an individual node s; we refer the reader back to Examples 4.2.1, 4.2.2 and 4.2.3 for clarity. To account for noise and for ease of interpretation, we present a city $s \in \mathcal{S}$ as belonging to a submarket r if and only if δ_{sr} is greater than 5 cents. Our results will be presented using this assumption for the rest of the thesis. When a region belongs to both submarkets in the sample, we superimpose the submarket indicator colors in the figure (i.e., blue over orange).

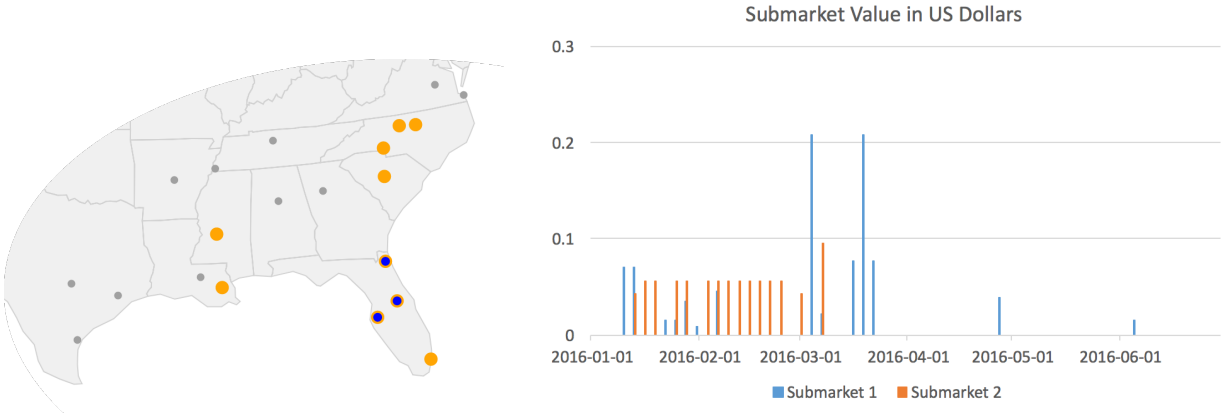

Figure 6.2: 2016H1 - No major submarkets detected except two independent peaks in late February and mid March.

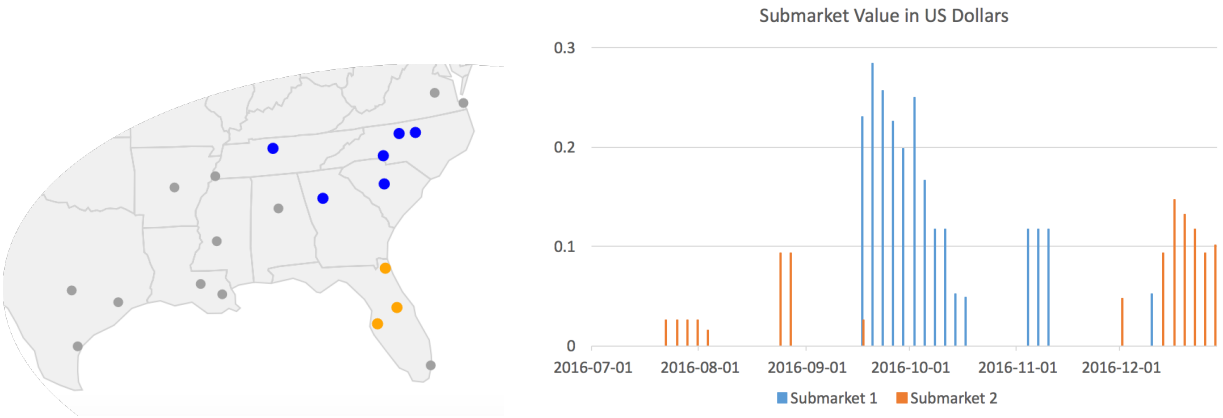

Figure 6.3: 2016H2 - A major submarket detected for eastern cities, which will be examined in detail below.

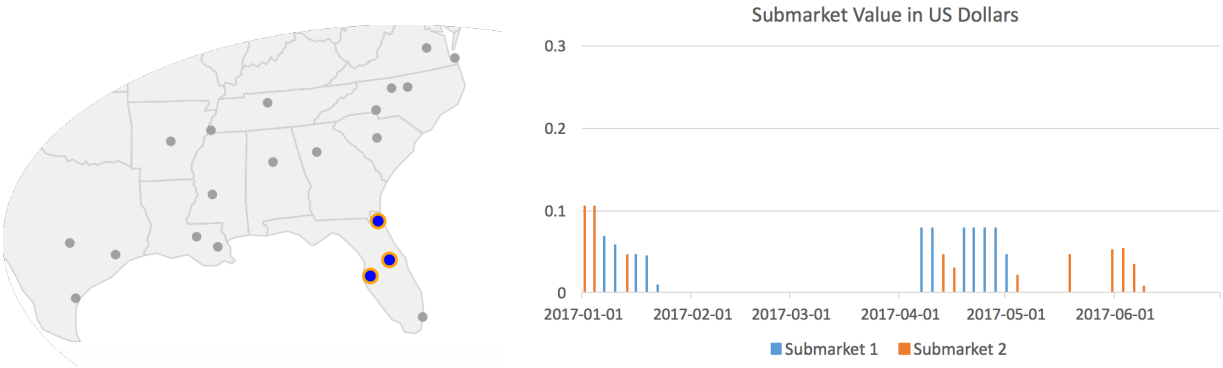

Figure 6.4: 2017H1 - No major submarkets, however, we note that the SDM detects a small submarket going into 2017H1, which is consistent (both in terms of regions and value) with the submarket at the end of 2016H2.

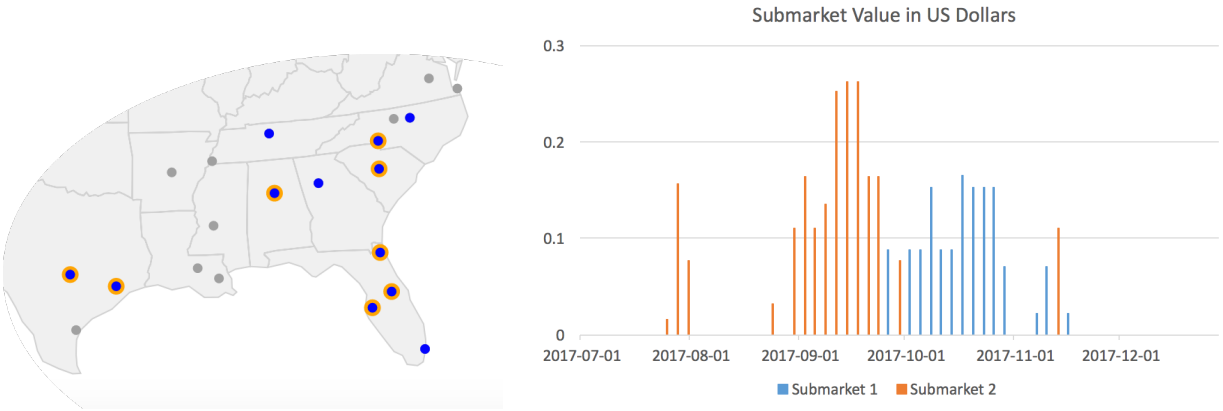

Figure 6.5: 2017H2 - Major submarkets detected affecting many regions of the US.

Period	V	V^o (SDM)
2016 H1	7.6	5.3
2016 H2	**14.1**	5.8
2017 H1	7.9	5.2
2017 H2	**10.0**	6.6

Table 6.1: The market integration value before apply the SDM (V), and after applying the SDM (V^o).

We begin the discussion by examining Table 6.1, which lists the V and V^o values. We observe that the periods of $2016H2$ and $2017H2$ have V values that are significantly higher than the periods of $2016H1$ and $2017H1$; this indicates that there is likely to be some market force (i.e., congestion) pulling apart prices during these periods. In contrast, we find that the values of V^o are relatively uniform across all samples. In particular, the value V^o after accounting for submarkets seems to be similar regardless of how small or large V was. This result is worth emphasizing. We refer to the idea of a neutral band existing between prices. This theory states that in the absence of congestion, prices differences may be randomly distributed but bound within this band. When there is congestion, the price differences will be pulled apart in a clean and "well-structured" manner. Thus, when there are bottleneck constraints in the underlying market, the SDM will be able to reduce the V (the maximum value of absolute price differences) significantly. However, when there is no congestion, prices can simply be randomly distributed within the band, and the SDM, which takes a specific structured approach to reducing error, is unlikely to capture any trend that can reduce the absolute value of price difference accross all regions. Thus, we believe that our results indicate that we have captured all major bottlenecks in the samples, and that the final values of V^o represent the region within which price differences can exist (and be randomly distributed).

We now discuss the results in more detail. The first observation we make is that there are specific regions that are consistently identified as being (or not being) in a submarket; we remind the reader that we define a submarket as set of regions that collectively suffer from a congestion surcharge, and a region that is not in a submarket simply represents a region in which we do not detect any abnormal price increases. We find that in all samples, the cities of Tampa, Orlando and Jacksonville in Florida are almost always identified as being in a submarket. This may indicate that there is some transportation friction in the delivery of gasoline to these cities. Interestingly, we find that Miami, a city at the southernmost corner of Florida, is rarely identified as belonging in a submarket. After investigating the transportation infrastructure in this area, we find that Miami obtains its gasoline from the refineries in Texas and Louisiana using tankers, whereas the cities further to the north obtain a portion of gasoline from pipelines. Our results seem to indicate that transportation through tankers may be more reliable and less prone to capacity constraints than pipelines (which seems in line with intuition - a disruption on a pipeline can have a major effect

on delivery whereas a disruption of a single marine vehicle is unlikely to disrupt delivery). On the other hand, we find that the cities in Texas and Louisiana are almost never identified as being in a submarket. This is consistent with what we would expect, since these cities are in immediate proximity with the refineries and are generally the first point of contact before gasoline is shipped further downstream to the eastern cities.

Our results also indicate that the 2016H2 and 2017H2 periods have the largest submarkets. These submarkets are temporally consistent with major disruptions that occured in this market during the sample time periods. In 2016H2, there were two pipeline disruptions in the pipeline network, and in 2017H2, Hurricane Harvey hit the Gulf Coast and caused the shutdown of significant portions of the transportation infrastructure. Since we will explore the results from 2016H2 in more detail below, we will only discuss the results for the 2017H2 sample. Our results in the 2017H2 sample show that the starting day of the detected submarket directly precedes the impact of Hurricane Harvey in the fall of 2017. When Hurricane Harvey hit the Gulf Coast, it caused significant disruptions in the supply and transportation systems, lasting several weeks; the numerous disruptions likely made prices very unpredictable. It is interesting to note that the V^o for period 2017H2 is higher than the V^o of 2016H2 despite the initial V value of 2017H2 being significantly lower. This may be explained by the fact that the price differences of 2016H2, despite being larger, were caused by a well-structured network disruption (i.e., an explosion at a single link) that the SDM was able to detect. On the other hand, the effects of Hurricane Harvey likely had more complicated impacts on the transportation network, generating price variations that are harder to capture with the SDM. We now take a more in-depth examination of our results from the 2016H2 sample.

6.1.3 Comparison with 2016 Pipeline Disruptions

In this section, we use our results from the 2016H2 period to provide a more in-depth analysis of the effect of two pipeline disruptions occuring in this period. We begin by describing the events.

The Colonial Pipeline is divided into two pipelines that run parallel to each other, referred to as Line 1 and Line 2. Line 1 of the Colonial pipeline ships gasoline, whereas the Line 2 typically ships heating oil, diesel and jet fuel. On September 9th, 2016, a major pipeline leak was discovered on Line 1 of the Colonial Pipeline in Shelby County (circled in red in Figure 6.6), and a partial shutdown of the pipeline immediately followed [27]. The shutdown lasted until September 21st, 2016, when the pipeline resumed full operating capacity. On October 31st, a deadly pipeline explosion in Shelby County again caused a partial shutdown of the pipeline, and operations were restarted on November 8th.These dates are labeled using a red marker in Figure 6.7.

We now examine submarket 1 from our 2016H2 sample. We first remind the reader that the results were obtained from the 2016H2 sample; in particular, these results were obtained from running the SDM on the prices of all the cities in the dataset, rather than just the cities highlighted

in the figure below. Nonetheless, the methodology was successful at identifying that the only cities that were significantly impacted are the cities directly downstream of the disrupted pipeline.

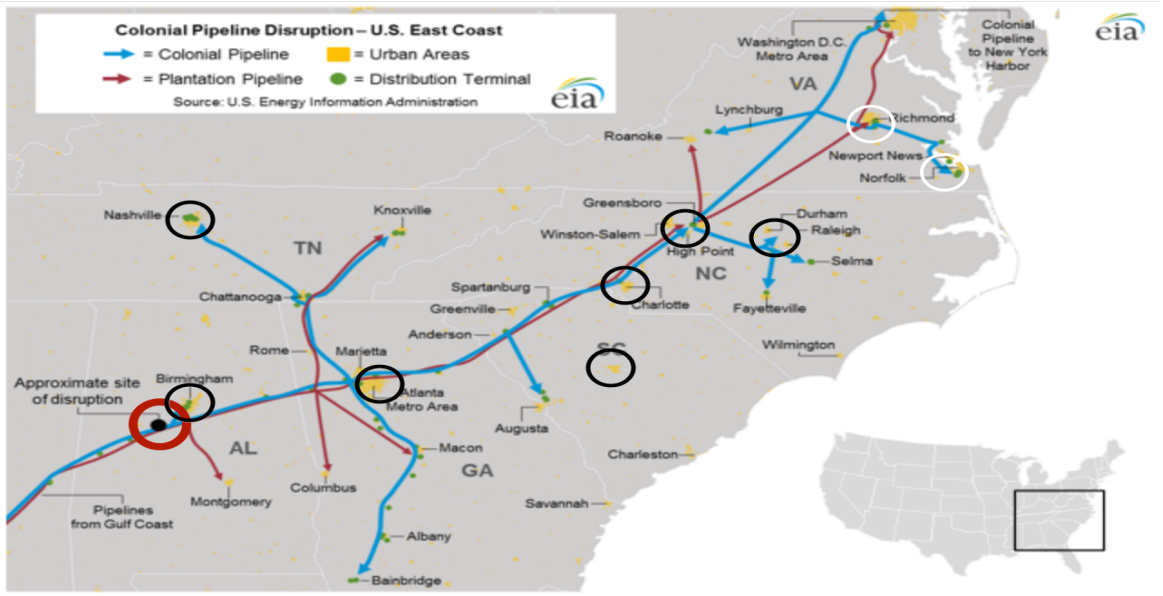

Figure 6.6: This figure presents the annotated map of the regions directly downstream of the disrupted pipeline. The regions (that we have data for) that belong to submarket 1 are highlighted in black, whereas those that do not are highlighted in white. The original source of the map is from: [1]

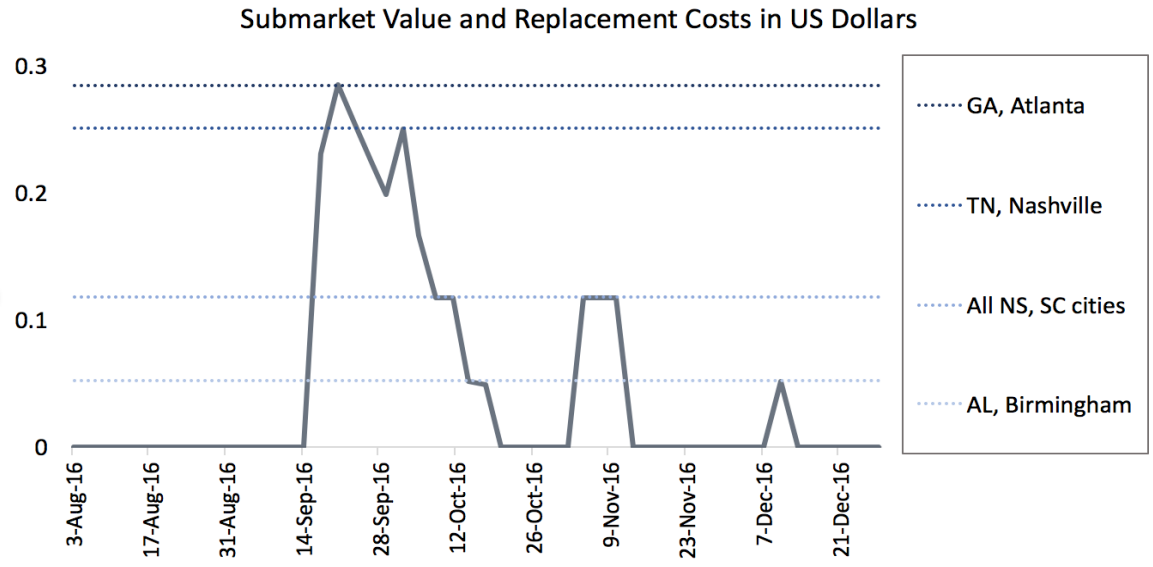

Figure 6.7: Submarket value and the replacement costs of submarket 1 in 2016H2. The actual dates of the disruptions are highlighted with the red marker.

We first examine the value of the submarket. The SDM detects a major submarket between Sept 17th and Oct 14th. The price surge is immediate and hits its peak shortly after the pipeline disruption. Gradually, over the next month, the price shock appears to subside. Even though the pipeline goes back to full operational capacity on September 26th, our results tell us that the submarket lasts for another 3 weeks, although with gradually decreasing magnitude. We suspect that this is likely due to the high demand for pipeline capacity following the disruption. In particular, many cities withdrew from local gasoline inventory and had to refill the depleted inventory following the disruption [10]. We also find that the SDM successfully identifies the second disruption; however, this second disruption appears to have had less of an impact in terms of both submarket value and duration.

We now discuss the geographical results that we obtain from the SDM. While the SDM was able to capture most of the cities that are directly downstream of the disrupted pipeline, the results also indicate that the cities farther downstream (i.e., Richmond and Virginia Beach, VA) are not part of the submarket. Interestingly, we find that there is a very plausible explanation for this. According to the EIA, there is a major junction point at Greensboro, NC, for both the Colonial and the Plantation Pipeline; a junction point is where gasoline can exit the pipeline. This might explain why cities downstream of this junction are not affected as much: there is still supply from the Plantation Pipeline. Furthermore, when Line 1 of the Colonial pipeline was disrupted, the EIA used Line 2 (which normally ships diesel, heating oil and jet fuel, rather than petroleum) to ship petroleum in order to alleviate the lack of supply in east coast cities. We hypothesize that these two factors combined may have helped the cities further downstream circumvent the effects of the pipeline disruptions.

We now discuss the δ_{sr} values obtained from our results. We remind the reader that the δ_{sr} values represent the replacement cost of each region s for r, i.e., the maximum degree that a region can be impacted by a congested link. Interestingly, we find that the replacement cost for Birmingham is low, which suggests that Birmingham did not experience a significant price surge despite being just a few miles away from the pipeline disruption. On the other hand, Nashville and Atlanta, the next two cities downstream of the disruption, appear to have the highest δ_{sr} values, which suggest that these regions experienced significant price increases as a result of the disruptions. Similarly, apart from Birmingham, there appears to be a distinct trend where the further downstream a region is from the disruption, the less of a price shock it experiences. This effect may be due to the layout of the transportation infrastructure, which suggests that the farther away a region is from a particular link, the easier it becomes to reroute gasoline to the region using the existing infrastructure.

Finally, we note that our results are obtained from running the SDM for a limited amount of time rather than to optimality. The results can thus be thought of as approximations from a heuristic rather than optimal solutions to the problem. We leave this issue for future work.

6.2 Validity of Results

The case study above shows that the results from the SDM are consistent geographically and temporally with the two documented network disruptions. Nonetheless, the SDM is built with the purpose of inferring the characteristics of all underlying constraints, including those that may be completely unobservable. For example, while our SDM was able to successfully identify the disruptions discussed in the case study, the results also indicate many smaller submarkets that cannot be associated with any major, media-reported event. In this case, validation of results naturally becomes a potential challenge. Nonetheless, there are certain features of the results that may suggest that the submarkets identified are real. First, we find that all regions identified in a submarket are all in spatial proximity; there are no major geographical gaps between these regions. This is generally what would be expected from a congested pipeline in this region, where the gasoline is flowing unidirectional from the Gulf Coast to the eastern cities. Secondly, we find that the SDM naturally identifies uninterrupted lengths of time of active submarkets (rather than short-lived sporadic time intervals that may be more indicative of detection of spurious noise). Finally, the values of the submarkets identified generally have a smooth unimodal trend consistent with what one would expect; prices will rise up until a point where the market forces begin to clear or alleviate the source of congestion. We believe that these characteristics of the detected submarkets can serve as strong indicators of reliable results.

Chapter 7

Conclusion

In this thesis, we first examine the relationship between a capacitated network and equilibrium prices in a competitive commodity market. We find that certain network structures, referred to as *structurally-integrated networks*, can guarantee a certain bound on price differences, referred to as the *neutral band*. We generalize the neutral band over the network and show that it is the tightest bound on price differences without making assumptions on consumer welfare and producer cost functions. We examine how prices are distributed when there is congestion in the system, and we derive a time series model of our prices where the connection between the transportation network and prices is directly observable. We then use these insights in deriving the submarket detection method, which allows us to estimate when submarkets occur, the magnitude of each submarket, and the degree to which different regions belong in the submarket. We used the SDM in a case study of the gasoline market and demonstrated the insights that can be gained from our approach.

We believe the results in this thesis can be applicable in a variety of different settings. From a prescriptive point of view, our results can aid policy makers in designing or evaluating markets where market participants trade and compete over a transportation network with limited capacity. In particular, we have discussed network qualities that can (or cannot) guarantee strong price integration. We believe these results can be used to design transportation networks that can facilitate and promote strong competition between market participants. This thesis also contributes to the descriptive side of analytics. We provide a well-structured way of analyzing prices generated from a market with transportation constraints. The methodology we propose provides a principled approach to estimating the effects of bottleneck constraints in competitive markets, and to determine how different regions may respond to a given transportation constraint. This enables us to learn various features of the underlying market. When we know that a resource is congested, the methodology can help to estimate the congestion surcharge caused by the congested transportation resource, and the degree to which different regions rely on the resource. On the other hand, when flow data is not known, the SDM can be used to identify regions that are likely to be suffering

from bottlenecks brought from day-to-day trading. We believe that this information can be used by policy-makers and regulators; for example, for capacity investment valuation, regulations on transportation access, or for security purposes (pipelines are consistently cited as a potential target of terrorist attacks [36]).

We believe that there are a few potential future directions for the work presented in this thesis; these extensions primarily revolve around improving the methodology. First, we believe that the SDM can be improved by using concepts recently developed in the field of anomaly detection. We believe it would be possible to use the ideas in this field to automate the selection of the input parameters of the SDM; in particular, to automate the selection of the submarket window length and the number of submarkets, both of which are currently input parameters. Secondly, we believe that the methodology can be made more robust for noisy datasets. Currently, the SDM relies on the assumption that the market being analyzed is competitive and that there would be strong price integration in the absence of congestion. However, it could be possible that the dataset contains a few regions that do not behave competitively and where prices are simply uncorrelated with the rest of the data. A potential solution could be the implementation of a pre-processing method, but we leave this as a future extension of the SDM. Finally, a third possible extension could be the combined use of pricing data and partial information of market characteristics. For example, while it might be difficult to obtain complete information on the entire southeastern gasoline transportation network, one might have access to the information of a subset of the network qualities. This information could then be used along with price data to obtain more accurate estimates on the effects of different congested links on regional prices.

Appendix

Tables and Supplemental Figures

Summary Statistics	Mean	Std	Min	Max	Range
AL, Birmingham	2.02	0.20	1.49	2.50	1.01
AR, Little Rock	2.04	0.18	1.50	2.40	0.91
FL, Jacksonville	2.19	0.20	1.68	2.71	1.03
FL, Orlando	2.19	0.21	1.64	2.71	1.07
FL, Tampa	2.19	0.21	1.67	2.73	1.06
FL, Miami	2.34	0.19	1.83	2.79	0.95
GA, Atlanta	2.32	0.21	1.77	2.90	1.14
LA, Baton Rouge	2.00	0.19	1.47	2.33	0.86
LA, New Orleans	2.06	0.19	1.52	2.41	0.89
MS, Jackson	2.01	0.18	1.50	2.44	0.94
NC, Charlotte	2.16	0.19	1.67	2.64	0.98
NC, Durham	2.20	0.18	1.74	2.62	0.88
NC, Greensboro	2.19	0.19	1.66	2.63	0.97
SC, Columbia	2.01	0.19	1.54	2.55	1.01
TN, Nashville	2.13	0.22	1.52	2.68	1.16
TN, Memphis	2.06	0.20	1.52	2.48	0.97
TX, Austin	2.05	0.21	1.50	2.59	1.09
TX, Houston	2.07	0.20	1.51	2.52	1.02
VA, Richmond	2.07	0.20	1.48	2.50	1.01
VA, Virginia Beach	2.07	0.19	1.52	2.52	1.00

Table 7.1: Summary statistics of daily prices from January 1st, 2016 to December 31st, 2017

City/Submarket	2016H1 S1	2016H1 S2	2016H2 S1	2016H2 S2	2017H1 S1	2017H1 S2	2017H2 S1	2017H2 S2
AL, Birmingham	0	0.04	0.05	0	0	0	0.15	0.15
AR, Little Rock	0	0	0	0.03	0	0	0	0
FL, Jacksonville	0.07	0.06	0	0.09	0.05	0.02	0.09	0.16
FL, Orlando	0.07	0.06	0.04	0.15	0.06	0.05	0.05	0.16
FL, Tampa	0.21	0.09	0	0.12	0.08	0.11	0.07	0.26
FL, Miami	0.03	0.06	0	0.03	0	0	0.15	0.02
GA, Atlanta	0	0	0.28	0	0	0	0.26	0.16
LA, Baton Rouge	0	0	0	0	0	0	0	0
LA, New Orleans	0	0.06	0	0.02	0	0	0	0
MS, Jackson	0	0.06	0.05	0	0	0	0	0
NC, Charlotte	0.02	0.06	0.12	0	0	0	0.08	0.11
NC, Durham	0.01	0.06	0.12	0.01	0	0	0.07	0
NC, Greensboro	0	0.06	0.12	0.03	0	0	0	0
SC, Columbia	0	0.06	0.12	0	0	0	0.09	0.14
TN, Nashville	0.05	0	0.25	0.04	0	0	0.26	0.1
TN, Memphis	0	0	0.05	0.04	0	0	0	0
TX, Austin	0	0.06	0	0.03	0	0	0.15	0.09
TX, Houston	0.03	0	0	0.03	0	0	0.07	0.08
VA, Richmond	0.05	0	0.05	0.05	0	0	0	0
VA, Virginia Beach	0	0	0.03	0.03	0.05	0.03	0	0

Table 7.2: The δ_{rs} values (i.e., estimated replacement costs) of each submarket for each region.

Bibliography

[1] U.S. Energy Information Administration. Major gasoline pipeline in southeast disrupted for second time in two months. https://www.eia.gov/todayinenergy/detail.php?id=28632, November 2016.

[2] Tarem Ahmed, Mark Coates, and Anukool Lakhina. Multivariate online anomaly detection using kernel recursive least squares. In *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*, pages 625–633. IEEE, 2007.

[3] Nathan S Balke and Thomas B Fomby. Threshold cointegration. *International economic review*, pages 627–645, 1997.

[4] Max Bennett and Yue Yuan. On the price spread of benchmark crude oils: A spatial price equilibrium model. Available at SSRN: https://ssrn.com/abstract=2894389, March 2017.

[5] Stephen PA Brown and Mine K Yücel. Deliverability and regional pricing in us natural gas markets. *Energy Economics*, 30(5):2441–2453, 2008.

[6] Stephen PA Brown and Mine K Yücel. What drives natural gas prices? *The Energy Journal*, pages 45–60, 2008.

[7] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):15, 2009.

[8] Haibin Cheng, Pang-Ning Tan, Christopher Potter, and Steven Klooster. Detection and characterization of anomalies in multivariate time series. In *Proceedings of the 2009 SIAM International Conference on Data Mining*, pages 413–424. SIAM, 2009.

[9] Shin C Chin, Asok Ray, and Venkatesh Rajagopalan. Symbolic time series analysis for anomaly detection: a comparative evaluation. *Signal Processing*, 85(9):1859–1868, 2005.

[10] Owen Comstock. Pipeline disruption leads to record gasoline stock changes in southeast, gulf coast. https://www.eia.gov/todayinenergy/detail.php?id=28172, September 2016.

[11] Helmuth Cremer, Farid Gasmi, and Jean-Jacques Laffont. Access to pipelines in competitive gas markets. *Journal of Regulatory Economics*, 24(1):5–33, 2003.

[12] Arthur De Vany and W David Walls. Pipeline access and market integration in the natural gas industry: Evidence from cointegration tests. *The Energy Journal*, pages 1–19, 1993.

[13] Caroline Dieckhöner, Stefan Lochner, and Dietmar Lindenberger. European natural gas infrastructure: the impact of market developments on gas flows and physical market integration. *Applied energy*, 102:994–1003, 2013.

[14] Mette Ejrnaes and Karl Gunnar Persson. Market integration and transport costs in france 1825–1903: a threshold error correction approach to the law of one price. *Explorations in economic history*, 37(2):149–173, 2000.

[15] Walter Enders and Pierre L Siklos. Cointegration and threshold adjustment. *Journal of Business & Economic Statistics*, 19(2):166–176, 2001.

[16] Jacob A Frenkel and Richard M Levich. Covered interest arbitrage: Unexploited profits? *Journal of Political Economy*, 83(2):325–338, 1975.

[17] Jacob A Frenkel and Richard M Levich. Transaction costs and interest arbitrage: Tranquil versus turbulent periods. *Journal of Political Economy*, 85(6):1209–1226, 1977.

[18] Steven A Gabriel, Shree Vikas, and David M Ribar. Measuring the influence of canadian carbon stabilization programs on natural gas exports to the united states via a 'bottom-up'intertemporal spatial price equilibrium model. *Energy Economics*, 22(5):497–525, 2000.

[19] Francesco Goletti, Raisuddin Ahmed, and Naser Farid. Structural determinants of market integration: The case of rice markets in bangladesh. *The Developing Economies*, 33(2):196–198, 1995.

[20] Barry K Goodwin and Nicholas E Piggott. Spatial market integration in the presence of threshold effects. *American Journal of Agricultural Economics*, 83(2):302–317, 2001.

[21] Clemens Haftendorn and Franziska Holz. Modeling and analysis of the international steam coal trade. *The Energy Journal*, pages 205–229, 2010.

[22] Patrick T Harker. Alternative models of spatial competition. *Operations Research*, 34(3):410–425, 1986.

[23] Patrick T Harker and Terry L Friesz. The use of equilibrium network models in logistics management: with application to the us coal industry. *Transportation Research Part B: Methodological*, 19(5):457–470, 1985.

[24] David F Hendry and Katarina Juselius. Explaining cointegration analysis: Part 1. *The Energy Journal*, pages 1–42, 2000.

[25] David F Hendry and Katarina Juselius. Explaining cointegration analysis: Part ii. *The Energy Journal*, pages 75–120, 2001.

[26] Mark J Holmes, Jesús Otero, and Theodore Panagiotidis. On the dynamics of gasoline market integration in the united states: Evidence from a pair-wise approach. *Energy Economics*, 36:503–510, 2013.

[27] ICF International. East coast and gulf coast transportation fuels markets. Technical report, U.S. Energy Information Administration, February 2016.

[28] Katarina Juselius et al. Testing structural hypotheses in a multivariate cointegration analysis of the ppp and the uip for uk. *Journal of econometrics*, 53(1-3):211–244, 1992.

[29] Eamonn Keogh, Jessica Lin, and Ada Fu. Hot sax: Efficiently finding the most unusual time series subsequence. In *Fifth IEEE International Conference on Data Mining (ICDM'05)*, pages 226–233, 2005.

[30] Ted G Lewis. *Critical infrastructure protection in homeland security: defending a networked nation*. John Wiley & Sons, 2014.

[31] Jessica Lin, Eamonn Keogh, Stefano Lonardi, and Bill Chiu. A symbolic representation of time series, with implications for streaming algorithms. In *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*, pages 2–11. ACM, 2003.

[32] Ming Chien Lo and Eric Zivot. Threshold cointegration and nonlinear adjustment to the law of one price. *Macroeconomic Dynamics*, 5(4):533–576, 2001.

[33] Stefan Lochner. Identification of congestion and valuation of transport infrastructures in the european natural gas market. *Energy*, 36(5):2483–2492, 2011.

[34] Vadim Marmer, Dmitry Shapiro, and Paul MacAvoy. Bottlenecks in regional markets for natural gas transmission services. *Energy Economics*, 29(1):37–45, 2007.

[35] Murthy Mudrageda and Frederic H Murphy. Or practice—an economic equilibrium model of the market for marine transportation services in petroleum products. *Operations Research*, 56(2):278–285, 2008.

[36] New Jersey Office of Homeland Security and Preparedness. Security considerations for pipeline systems. https://www.njhomelandsecurity.gov/analysis/rb-pipelines, April 2016.

[37] Matthew E Oliver, Charles F Mason, and David Finnoff. Pipeline congestion and basis differentials. *Journal of Regulatory Economics*, 46(3):261–291, 2014.

[38] Haesun Park, James W Mjelde, and David A Bessler. Time-varying threshold cointegration and the law of one price. *Applied Economics*, 39(9):1091–1105, 2007.

[39] Rodney J Paul, Dragan Miljkovic, and Viju Ipe. Market integration in us gasoline markets. *Applied Economics*, 33(10):1335–1340, 2001.

[40] H-E Reimers. Comparisons of tests for multivariate cointegration. *Statistical papers*, 33(1):335–359, 1992.

[41] Paul A Samuelson. Spatial price equilibrium and linear programming. *The American economic review*, 42(3):283–303, 1952.

[42] Nicola Secomandi. On the pricing of natural gas pipeline capacity. *Manufacturing & Service Operations Management*, 12(3):393–408, 2010.

[43] Alison Sider and Nicole Friedman. Gas prices jump after pipeline fire. https://www.wsj.com/articles/gas-diesel-prices-spike-following-alabama-pipeline-fire-1478010889, November 2016.

[44] Jesse Snyder. Oil firms resume rail shipments as crude oil pipelines fill up again. https://business.financialpost.com/commodities/energy/oil-firms-resume-rail-shipments-as-pipelines-fill-up-again, February 2017.

[45] David I Stern. A multivariate cointegration analysis of the role of energy in the us macroeconomy. *Energy economics*, 22(2):267–283, 2000.

[46] Farrukh Suvankulov, Marco Chi Keung Lau, and Fatma Ogucu. Price regulation and relative price convergence: Evidence from the retail gasoline market in canada. *Energy Policy*, 40:325–334, 2012.

[47] Takashi Takayama and George G Judge. An intertemporal price equilibrium model. *Journal of Farm Economics*, 46(2):477–484, 1964.

[48] Takashi Takayama and George G Judge. *Spatial and temporal price and allocation models*. North-Holland, 1971.