A Cournot Framework for Supply Chain Analysis
With Application to the Net Neutrality Debate

Working Paper

Paul Laskowski[1,2] Pedro Ferreira[2,3]

1 – School of Information, UC Berkeley
2 – Instituto Superior Tecnico, Technical University of Lisbon, Portugal
3 – Engineering and Public Policy Department and Heinz College, CMU

Abstract
In the years since net neutrality first emerged as a nexus of policy debate, a substantial literature has emerged to model possible departures from today's industry practices (see Cheng, et al. 2011, Choi and Kim 2010, Economides and Tåg 2009, Hermalin and Katz 2007, Musacchio, et al. 2009). While theoretical approaches vary, each study is ultimately rooted in the classic building blocks of Bertrand and Cournot competition. This common foundation has yielded rapid gains to our understanding, but it is also related to at least three broad challenges. First, as competition games are layered to depict various network exchanges, tractability is strained. Second, studies lose generality as extra assumptions are added to capture the market power of network providers. Finally, the term neutrality obscures a substantial variety of alternatives to today's industry. Most existing models can only compare two distinct regimes, making it hard to develop a holistic understanding of the larger neutrality space.

Believing that these limitations are rooted in the very foundation of Bertrand and Cournot competition, we consider a novel, and radical, question: can these building blocks be adapted to better target the needs of network research? Our approach generalizes Cournot competition in a way that is fully consistent with existing theory, but provides a natural extension from a single market to more complex network environments. To represent various neutrality regimes, we begin by formalizing the concept of a supply chain. This paradigm is both simple, abstracting the interaction of network players, and flexible, capturing key differences between regimes. Our main result extends Cournot competition to this setting, allowing us to compute prices at each stage of production. Moreover, our method is highly tractable, transparently captures market power, and treats all technologies symmetrically, isolating the effects of market structure.

Because of the flexibility of our theory, we may represent an unprecedented number of neutrality regimes in a universal model. In preliminary results, we compare three distinct regimes, identifying a number of effects that have not been previously highlighted in the literature. If neutrality is replaced by a regime of *passage fees* – in which service providers must pay to access each network provider's customers – a double marginalization effect dramatically increases prices and lowers welfare. On the other hand, we consider a *duopoly-price rule* in which two network providers compete to sell transit to service providers. This regime gives network providers nearly as much profit as the duopoly price rule, or even more in some cases, but without the dramatic welfare-loss. Our results begin to paint a more comprehensive view of neutrality. In future work, we will study how each regime affects investment in both infrastructure and services.

1 Introduction

In the years since net neutrality first emerged as a nexus of policy debate, a substantial literature has emerged to model possible departures from today's industry practices (see Cheng, et al. 2011, Choi and Kim 2010, Economides and Tåg 2009, Hermalin and Katz 2007, Musacchio, et al. 2009). While theoretical approaches vary, each study is ultimately rooted in the classic building blocks of Bertrand and Cournot competition. This common foundation has enabled researchers to build a number of successful models in just a few years. At the same time, these competition games can be related to at least three broad challenges that pervade the neutrality field.

First, network users interact with network providers, service and content providers alike. As Bertrand and Cournot games are layered to depict these exchanges, tractability quickly becomes an issue. Complex expressions make it difficult to interpret results, constraining the range of assumptions that can be explored. This both limits the realism of a model and makes it difficult to establish robustness.

Second, network providers exercise considerable market power in the industry, elevating prices well above costs. In practice, capturing this phenomenon with standard competition games involves tradeoffs. To prevent network providers from competing prices away to marginal cost, studies must employ extra assumptions, such as Hotelling differentiation (Hermalin and Katz 2007, Economides and Tag 2009) and captive consumers (Musacchio et al 2009). Justifying a specific assumption about the relationship between network firms may be difficult. Furthermore, there may be no way to know whether a result arises from market structure, or from the different treatment of network and service firms. As a result, models lose generality and flexibility.

Finally, and most importantly, the term neutrality obscures a substantial variety of alternatives to today's regime. Various studies reveal several dimensions contained within the larger debate. These include, for example, the direction of payments among network parties, the presence of multiple service classes, and the degree of discrimination within each service class. Rather than a binary choice, in other words, we face a more general *neutrality space*. Most existing models, however, can only compare two distinct regimes, and differing assumptions make it hard to develop a holistic understanding of the larger space.

These limitations – related to *tractability*, *realism*, and *scope* – are not general problems of Bertrand and Cournot competition, but only of applying these games, which are well-tailored to traditional markets, to the network environment. At the same time, the near-universality of these games makes them all-but-impossible to avoid when writing a model to study neutrality. In response, this study explores a novel approach. Instead of rearranging Bertrand and Cournot games into further models, we will scrutinize this very foundation, and ask whether traditional competition games can be adapted to better target the needs of network research.

The first aim of this study, in other words, is to design a general purpose tool for modeling network environments. While our framework is novel, it is also firmly grounded in the classic Cournot model. Using a change of decision variables, we are able to provide a natural extension from a single market to more complex network environments. While this approach may seem unorthodox, our framework is fully consistent with existing theory, and behaves naturally in a variety of familiar scenarios.

Before we can use our framework to compare neutrality regimes, each one must be encoded using a common set of primitives. Such a language, or meta-framework, must be tractable, abstracting the interaction of network actors, as well as expressive, capturing the key differences between regimes. Our approach balances these aims using the paradigm of a supply chain. This is a powerful tool for studying neutrality, allowing payments to flow in arbitrary configurations among network actors.

As we will demonstrate, many different regimes can be described in terms of a supply chain, highlighting the goods that are exchanged between network participants. The links in the supply chain express how goods and services are sold from firm to firm, and their relationship to each other. Moreover, such a model encapsulates key differences between regimes in a compact format, allowing a simple, graphical comparison.

On top of our supply chain model, we will develop a generalized theory of Cournot competition. By employing quantity competition throughout a supply chain, our framework will overcome the three major limitations of the neutrality field that we described above. Because there are multiple ways to distribute profits across a supply chain for a given demand profile, we will develop axioms to identify unique prices at every point in the chain.

Because of the flexibility of our theory, we may represent an unprecedented number of neutrality regimes in a universal model. In section 4, we will apply our theory to compare three distinct regimes, demonstrating a number of effects that have not been previously highlighted in the literature:

Zero-Price Rule: We model a scenario in which service providers do not exchange money with network providers – a benchmark rule commonly used to represent neutrality in the literature. Our formulation is unique in that networks and services are treated symmetrically, so results are driven by solely by differences in market concentration and cost of entry.

Uniform Passage Fees: We consider a framework in which service providers must pay network providers to access their customers, whether they are directly connected or not. We call such payments “passage fees” and discover a double-marginalization effect that dramatically increases prices and reduces welfare compared to the zero-price rule. Passage fees may be uniform, or we may adapt our framework to allow discrimination among specific service providers or classes of service.

Duopoly-Price Rule: As an alternative to passage fees, we ask whether access networks would want to leave their peering agreements with other backbone providers and sell them transit instead. Transit prices would then be constrained by duopoly competition between access providers. We are surprised to find that this scenario provides network providers with nearly as much profit as uniform passage fees (and in some cases more profit), while maintaining nearly as much welfare as the zero-price rule.

Taken together, these cases provide a newly comprehensive view into net neutrality, delineating alternate regimes and highlighting their pros and cons. In future work, we will study how each regime affects investment in both infrastructure and services.

2 Neutrality Regimes in the Economic Literature

Within economics, three main lineages bear on the topic of net neutrality. Not only do the methods vary between these, so do the actual regimes that they target for comparison. At least three industry scenarios can be identified in the literature, with each pair of these the subject of a separate economic lineage.

Representing a neutral network, most studies employ what become known as the **zero-price rule** (Economides & Tag 2009, Musacchio et al. 2009, Kramer & Wiewiorra 2010, Cheng et al. 2011, Choi & Kim 2010). The common feature underlying these models is that network providers exchange payments solely with end-users. Service providers pay nothing (hence, a zero-price) to connect to the network and communicate with their customers.

The zero-price rule is a tremendously popular, though imperfect, depiction of today's industry. To be more precise, a service provider must enter a contract with some ISP in order to access the internet. In this context, the network provider is said to sell *transit* to the service provider. This good includes the right to send data to any destination on the internet, so a service provider need only purchase transit from a single ISP. Although service providers pay for transit, competition among providers has long driven down prices (Musacchio et al. 2009). In fact, transit prices fell an average of 61% each year between 1998 and 2010 (Norton 2010). The zero-price rule can be viewed as an extrapolation of this trend.

One strand of papers that uses the zero-price rule as a baseline arises from the classic economic model of two-sided markets. In this framework, a network provider faces demand from consumers on one side, and demand from service providers on the other. Under the zero-price rule, the network provider can only charge a price from consumers. Under the non-neutral regime, the network provider can charge a separate fee against service providers.

As mentioned above, service providers already pay network providers today, in the form of transit fees. A key feature of the two-sided market studies, however, is that service providers must pay every network provider separately to communicate with their customers (Economides & Tag 2009, Musacchio et al. 2009). Hence, unlike transit, which is the right to send data anywhere on the internet, network providers are selling just the right to reach their own customers. In order to distinguish this right from transit, throughout this study, we will refer to it as *passage*. Studies in the two-sided market lineage typically restrict network providers to choosing a single price for passage. We will therefore refer to this scenario as a **uniform passage fee** regime.

Economides and Tag present a two-sided market model that allows a duopoly of network providers (2009). They argue that uniform passage fees are welfare-reducing compared to the zero-price rule, though consumers face lower prices and are better off. Musacchio, et al. present a model in which N network providers can charge a uniform passage fee (2009). They find that the welfare effects of this regime are ambiguous compared to the zero-price rule. While the authors model more than one network provider, each one is assumed to enjoy a monopoly over their customer base, so the model is not targeted to consider what effect passage fees have on oligopolistic competition.

Moving in another direction, a separate lineage of studies similarly begins with a zero-price regime, but instead of introducing a single passage fee, imagines dividing the network into a base tier and a higher priority tier of service. The relative qualities of each tier are then modeled using queueing theory. Typically, the base tier remains free, but service providers must pay a passage fee in order to transmit through the higher tier. We will label this scenario **multiple service classes**.

In this vein, Kramer and Wiewiorra consider a monopolist network provider that can sell prioritized access to a continuum of content providers (2010). They find the allowing such discrimination increases short-term welfare, as well as incentives to invest in network capacity and content innovation.

Cheng et al. consider a scenario with two content providers. Under multiple service classes, the network provider is allowed to sell priority access to one or both content providers, prioritizing their packets (2011). They find that allowing such charges either increases welfare and consumer surplus, or leaves them unchanged.

Choi and Kim similarly consider two content providers that use a monopolist network provider (2010). Under multiple service classes, the network provider can sell a priority right to one content provider, and the authors argue that this increases consumer surplus. The effects on short-run welfare are ambiguous. Incentives to invest in network capacity can be lower in the multiple service classes regime, because more capacity reduces the quality difference between service classes.

Completing the circle, a third strand of the literature can be said to compare the uniform passage fee and multiple service classes regimes. Hermalin and Katz model a network provider in a framework of product-line restrictions (2007). The authors consider whether a planner would want to restrict the network provider to selling a single product, effectively creating a uniform passage fee scenario. They find that the welfare effects of such product restrictions are ambiguous, but suggest that such restrictions will tend to be welfare-reducing in practice. Unlike the previous studies mentioned, they do not model the zero-price rule.

Figure 1 depicts the relationship of the three main lineages within the neutrality literature, and the three regimes they investigate.

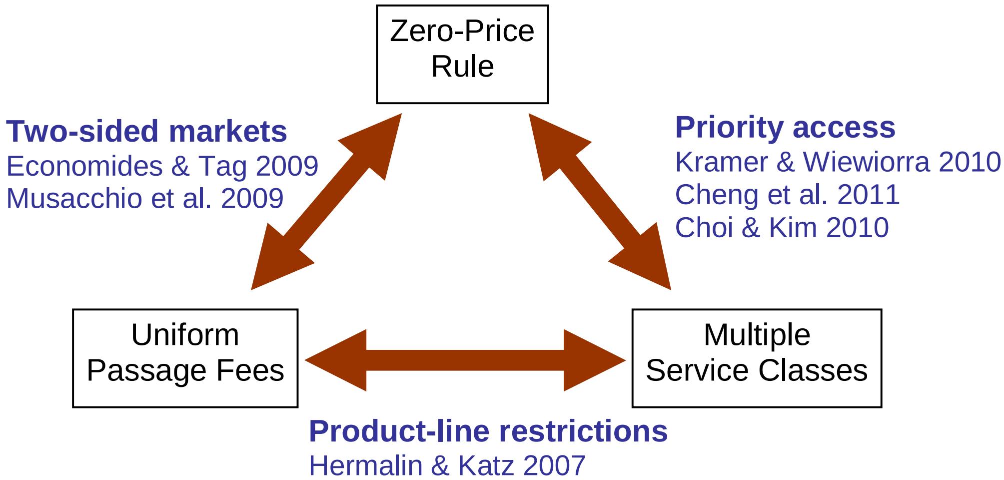


Figure 1: Major Net Neutrality Lineages

3 A Cournot Theory of Supply Chain Behavior

The previous section described three neutrality regimes found in the economic literature. As we have seen, researchers have been successful in writing models to compare pairs of these, but the disparate assumptions they employ make it difficult to compare results across lineages, or to gain a broader understanding of the neutrality space. In contrast to these studies, we will construct our own modeling foundation, extending Cournot competition to a general supply-chain setting. This work bears some resemblance to Salinger's model of two vertically-arranged markets (1988). Unlike his study, which is limited to a fixed arrangement of two exchanges, our model allows an arbitrary arrangement of exchanges within a supply chain. Moreover Salinger's technique requires firms in the upstream market to select their actions first, while our framework allows any timing of player decisions.

Our work also builds upon a previous model by Laskowski and Chuang, in which Cournot competition is generalized to allow both network providers and service providers to select quantities to sell to a user market (2010). The current study further generalizes Cournot competition to the more general setting of a supply chain.

As we will detail below, our model of a supply chain corresponds closely to the traditional notion for manufactured products – goods are sold from firm to firm and assembled with other goods before being sold to consumers. As we apply this framework to network neutrality, there are two features that may cause some concern, and require us to update our intuition. First, many goods that we wish to model in the network environment are intangible. To capture the core interactions relevant to neutrality, we will need to include services and contractual obligations, such as transit, as goods in the chain.

Second, network goods are typically not assembled in a physical sense. For example, a service provider needs to buy transit before it can sell service to an end-user, but we do not normally think of transit as a part of the service product. To resolve this issue, we will take a broad view of assembly to include any instance in which an input is required for a good to continue along the supply chain. Thus, assembly denotes any requirement for a specific input, however it is used.

The supply chain paradigm compels us to focus on the goods exchanged by network players, and their relationship to each other. Thinking in terms of goods, instead of, say, contracts and fees, may take some getting used to. Given a new type of fee, for example, we must ask what it is that a network player is paying for, and encode that as a good within the supply chain. That good must then be assembled with other goods to signify where the fee is required. As we will see, however, this approach is flexible enough to capture a great variety of neutrality regimes. Moreover, we will be able to express the differences between different types of fees in a clear, graphical format.

Once we formalize our notion of a supply chain in the next subsection, we will be ready to extend Cournot competition to run over this foundation. Our use of Cournot competition provides a number of advantages over existing techniques. First, it is mathematically elegant. Expressions remain tractable, even as the supply chain grows to capture complex market scenarios. We may also alter game timings, or add extra assumptions, for greater expressive power.

Second, Cournot competition captures market power in a simple and transparent way. To model a network duopoly, models based on price competition must employ extra assumptions so that network providers do not compete away prices to marginal cost. Hermalin and Katz (2007), and Economides and Tag (2009) differentiate network providers on Hoteling's line. Musacchio et al (2009) assume that consumers are held captive to a specific network provider. Cournot competition similarly ensures that network providers make a profit, but without specific assumptions about how the technologies are related, or how users select a network. By abstracting away from such details, the model remains more general.

Finally, our framework uniquely allows networks and services to be treated symmetrically; it is entirely technology-agnostic. This means that the distribution of profits is driven entirely by market structure, and is never an artifact of treating networks and services differently. Put another way, our framework allows us to isolate the effects of the number of each type of firm or costs of entry, creating a valuable baseline for analysis.

We will begin our treatment in the next subsection by defining a supply chain in terms of simple graph primitives. Section 3.2 discusses Cournot competition in the context of a single market. This will give us the chance to introduce notation and build the intuition we need for the general model. In section 3.3, we will turn our attention to full supply chains, and sketch our most powerful results. Finally, in section 3.4, we will use our main theorem to formally define a Cournot game. Full technical details may be found in the appendix.

3.1 The Network as a Supply Chain

Mathematically, we define a *supply chain*, G, as a directed acyclic graph with nodes V and edges E, containing

1. A set of source nodes, $V_I \subset V$, which we call *inputs* or *input goods*.
2. A set of non-source nodes, consisting of a set of *assembly nodes*, $V_A \subset V$ and a set of *market nodes*, $V_M \subset V$, which we depict with a triangle symbol. We will call the incoming edges of each market node *market edges*, $E_M \subset E$.
3. A unique sink, which may be either an assembly node or a market node, and which we call the *consumer market*, $C \in V$.

Abusing notation slightly, we will also use the variable denoting a graph to represent the set of all elements of the graph, so $G = V \cup E$. For each node, n, we will write $\mathrm{inc}(n) \in E$ for its incoming edges. Given a supply chain, G, we define a final product, d, as an induced subgraph of G, with the following properties:

1. d includes the consumer market, and no other sinks.
2. If an assembly node, a, is in d, $\mathrm{inc}(a) \subset d$.
3. If a market node, m, is in d, exactly one edge in $\mathrm{inc}(m)$ is in d.

Let D be the set of all final products, and $H = \{h : D \to \mathbb{R}\}$ be the vector space of real final product quantities. Let h_d be the unit vector that assigns a quantity of 1 to final product d.

We will also want to describe the flow of goods along the supply chain, from input to final product. To do this, we may define a *quantity flow*, f, as an assignment of a quantity to every node and edge in G, $f : G \to \mathbb{R}$, such that,

1. The quantity of each node except the consumer market equals the total quantity of all outgoing edges.
2. The quantity of a market node equals the total quantity of all incoming edges.
3. The quantity of an assembly node equals the quantity of each individual incoming edge.

For mathematical convenience, we will allow the quantities in a quantity flow to take on negative values. Let F be the set of all quantity flows. It is easy to check that F is a vector space where addition and scalar multiplication are over each component. The following lemma is proved in the appendix.

Lemma 1. F has dimension $|E_M| - |V_M| + 1$.

For every set of final product quantities, there is a natural quantity flow that can be said to produce those quantities. Specifically, given a vector of final product quantities, $h \in H$, we may define $\phi(h)$ as the quantity flow that assigns to each component, $g \in G$, the total quantity of final products that include that component:

$$\phi(h)(g) = \sum_{d \in D | g \in D} h(d), \text{ for all } g \in G \tag{1}$$

In particular, we will let $u_d = \phi(h_d)$ be the flow that produces a unit quantity of final product *d*. It is important to note that in general, ϕ is not injective. This means that given a set of final product quantities, we can derive the unique quantity flow that produces those quantities, but a single quantity flow could be associated with more than one vector of final product quantities.

3.2 A single Cournot market

Before presenting our general Cournot framework, we will begin in this subsection with just a single market node. This scenario corresponds to the classic Cournot model, and we will take the opportunity to introduce our own notation. This will give us the intuition we need to extend the game over an entire supply chain.

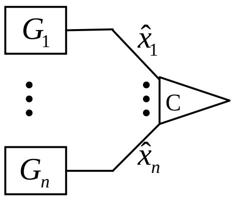
In the classic Cournot setup, *n* firms produce imperfect substitute products and choose quantities. Prices are determined by an inverse demand function, $\mathbf{t} = \{t_i\}$, where $t_i(x_1,...,x_n)$ is the price of firm *i*'s product, and x_i is the quantity firm *i* produces.. We will assume that there is a maximum total quantity, *N*, such that the total profit of all firms may be non-negative.

So far, we have given a more-or-less standard description of Cournot competition, focusing on how much of the market the firms supply. For our purposes, however, we will find it profitable to focus instead on how much of the market the firms decide to leave unsupplied. Define the quantity restriction for firm *i*, $\hat{x}_i$, to be how many fewer customers the firm supplies than it would have if all customers were divided equally: $\hat{x}_i = N / n - x_i$. This form is chosen so that the total quantity restriction is the number of unsupplied consumers:

$$\sum \hat{x}_i = N - \sum x_i \,. \tag{2}$$

Notice that quantity restrictions may be negative, so there is a one-to-one correspondence between quantities and quantity restrictions. This means that our game is fully equivalent to regular Cournot competition in a mathematical sense. The change, in other words, is merely one of emphasis, highlighting the total number of consumers as a reference point. While our change of decision variables has no effect on a single market, it does provide an intuitive way to extend Cournot competition over a supply chain.

We will represent this basic Cournot market graphically with the triangle symbol, below. In this graph, which we will explain in the next section, the lines represent each good, G_i, entering the market, with the corresponding quantity restriction labeled alongside.



If our change of decision variables strikes the reader as somewhat artificial, rest assured that it may be considered temporary. Quantity restrictions will give us a natural way to join Cournot games together to describe complex supply chains. In the end, though, we will see that firm's strategy can be seen as a solution to a regular Cournot game, and our framework may be rephrased in terms of quantities.

3.3 General Model

Building on the previous subsection, we are now ready to extend Cournot competition to an entire supply chain. Our procedure involves a number of steps. Very briefly, we must first specify how a quantity restriction reduces the quantity flow in different parts of a supply chain. These reductions are summed together for all quantity restrictions to find the resulting flow. Because restrictions are relative to a maximum amount, there is never a mismatch between decisions in different parts of the chain. When firms in one market node reduce their output, for example, firms in other parts of the supply chain automatically reduce their outputs to compensate.

After computing quantities, we are ready to compute prices in each market node. This is straightforward for a single market, since there is a unique set of prices that fulfills demand. Things become more complicated in a supply chain, however, since prices can be increased in one node and decreased in other nodes without changing the prices of the final products. We therefore take an axiomatic approach, adding further assumptions to constrain the set of possible prices. Our central theorem uses four assumptions to identify a unique set of prices at each stage of production. This will allow us to use our framework in much the same way as traditional Cournot competition to construct games.

Consumers are represented by an inverse demand function, $\mathbf{r} = \{r_d\}_{d \in D}, r_d : H \rightarrow \mathbb{R}$, where $r_d(h)$ represents the price of final product *d* as a function of final product quantities *h*.

Because firms in our model will make decisions at the granularity of quantity flows, and cannot in general control the quantity of individual final products, we will assume that demand only depends on the quantity flow. Specifically, given two vectors of final product quantities, $h_1, h_2 \in H$, produced by the same quantity flow, $\phi(h_1) = \phi(h_2)$, we assume

a) The resulting price vectors are equal, $\mathbf{r}(h_1) = \mathbf{r}(h_2)$.
b) The resulting total revenues are equal, $\sum_{d\in D} h_1(d) r_d(h_1) = \sum_{d\in D} h_2(d) r_d(h_2)$.

According to the next lemma, which is proved in the appendix, these assumptions let us describe demand in terms of prices exerted at each market edge:

Lemma 2: The inverse demand can be expressed in terms of a price exerted at each market edge, where the price of each final product is the total of the prices charged at its edges: There exists $\mathbf{t} = \{t_e\}_{e\in E_M}$, $t_e : F \to \mathbb{R}$, such that $r_d(h) = \sum_{e\in d\cap E_M} t_e\left(\phi(h)\right)$ for all $d \in D, h \in H$.

Note that in general, the market-edge prices, $\{t_e\}_{e\in E_M}$, are not unique (in the proof of the lemma, the choice of θ^{-1} is not unique). For example, it may be possible to increase the prices in one market while decreasing the prices in another market by the same amount, without affecting the price of any final products. Nevertheless, we will usually specify demand in terms of market-edge prices, bearing in mind that multiple specifications may be equivalent.

Analogously to the single stage game, we will have firms choose quantity restrictions, and then use these to compute the quantity flow. To every market edge, $e \in E_M$, we associate a quantity restriction, $\hat{x}_e$, and write $\hat{\mathbf{x}}$ for the vector of all such restrictions. We would like each restriction to decrease the quantity flow at its specific edge, but in order to do this, it may be necessary to reduce the flow along other edges to maintain a valid quantity flow. In general, there are many ways that a restriction at one edge can affect other parts of a supply chain, [does any linear map automatically work with theorem 1?] and our chief criterion is that this process be intuitive. Our method will be based on a simple rule: in essence, a quantity restriction will be split evenly each time the supply chain branches.

To achieve this, we first define the *upstream restriction*, $\hat{w}_c$, for each supply chain component $c \in G$, recursively as follows:

1. For any node, $v \in V$, $\hat{w}(v)$ is the total of the upstream restrictions of all incoming edges.
2. For any edge, $e \in E$, $\hat{w}(c)$ is the upstream restriction of its parent node, divided by the number of outgoing edges that parent has, and plus the edge's own quantity restriction, $\hat{x}_e$, if *e* is a market edge.

We assume that it is the upstream restrictions that determine the relative flow of each input to a market node. First, we assume that there is a maximum magnitude of quantity flow, *N*, such that the total profit of all final products is non-negative,

$N = \max_{f \in F}\left(|f| \,\middle|\, \sum_{d \in D} x_d t_d(d) \geq 0\right)$. Next, we define the quantity flow induced by $\hat{\mathbf{x}}$, $f \in F$, as the unique quantity flow such that,

1. $f(C) = N - \hat{w}_C$
2. For any two incoming edges of the same market node, *a* and *b*, $f(a) - f(b) = \hat{w}_b - \hat{w}_a$. Equivalently, for each market node, $m \in V_M$, there exists a constant c_m such that if *e* is an incoming edge of *m*, $f(e) = c_m - \hat{w}_e$.

Let $\omega : \mathbb{R}^{|V_M|} \to F$ be the map that takes each restriction vector to the quantity flow it induces. It is easy to see that the $\{\hat{w}_e\}$ are linear in $\hat{\mathbf{x}}$, and *f* is affine in the $\{\hat{w}_e\}$, hence ω is an affine transformation. In general, we will see that ω is surjective but not one-to-one. The following lemma will characterize the set of restrictions that result in the same flow.

Lemma 3: Given any flow, $f \in F$, and a set of total restrictions for each market, $\hat{s}_m, m \in V_M$, with $\sum \hat{s}_m = N - f(C)$, there exists a unique restriction vector, $\hat{\mathbf{x}}$, such that $\sum_{e \in \mathrm{inc}(m)} \hat{x}_e = \hat{s}_m$ for all $m' \neq m$ and $\omega(\hat{\mathbf{x}}) = f$. These vectors make up the inverse image, $\omega^{-1}(f)$.

A proof is given in the appendix. To represent the induced quality flow at a market edge, *e*, we will use the corresponding quantity restriction variable without the hat, $x_e = f(e)$, with $\mathbf{x} = \{x_e\}_{e \in E_M}$. Having computed quantities, we are ready to compute prices at each market node. Mirroring our representation of demand, we will find it convenient to work with marginal prices [is that the right term for it? I mean the markup over the price paid to acquire the good]. For each market edge, $e \in E_M$, let the marginal price at that edge be $\bar{x}_e$, represented by a bar over the corresponding restriction variable, with $\bar{\mathbf{x}} = \{\bar{x}_e\}_{e \in E_M}$. The total price of a final product, *d*, is then the sum of the marginal prices at each of its market edges, $\sum_{e \in d \cap E_M} \bar{x}_e$.

In classic Cournot competition, there is a unique set of prices such that demand is fulfilled and consumers are indifferent between purchasing all products. Unfortunately, while the above lemma assures us that there is a set of market-edge prices that fulfill demand, it will not, in general be unique. It may be possible, for example, to subtract some amount of money from the prices at one market node, and add it to the prices at another market node while preserving all final product prices.

Because of this, we will need more assumptions to identify the most natural set of market-edge prices, as a function of quantity restrictions. Our approach will use the following four:

1: Demand is fulfilled, $\overline{\mathbf{x}} \cdot u_d = \mathbf{t} \cdot u_d$ for all $d \in D$.
2: When the domain is restricted to a fixed quantity flow, *f*, prices are linear in quantity restrictions, $\hat{\mathbf{x}}$.
3: If the total restriction at any market node is zero, changing the restrictions at that node while maintaining the total quantity restriction does not affect the sum of the revenues at all other market nodes.
4: If the total restriction at any market node is zero, and the restrictions at all other nodes are fixed, the maximum total profit at that node is zero. That is, for any market node $m \in V_M$, with edges $E_m \in E_M$, given restrictions at all other market nodes, $\{\hat{x}_e^0\}_{e \in E_M / E_m}$, $\max_{\hat{\mathbf{x}} | \hat{x}_e = \hat{x}_e^0,\, e \in E_M / E_m} \sum x_e \overline{x}_e = 0$.

Theorem 1: Given a set of quantity restrictions, $\hat{\mathbf{x}}$, with non-zero total, $\sum \hat{\mathbf{x}}_e \neq 0$, such that the total flow at each market node is non-zero, there exists a unique price function, $\overline{\mathbf{x}} : \mathrm{E}_M \to \mathrm{E}_M$, such that assumptions 1-4 are fulfilled.

The complete proof is given in the appendix.

3.4 Game Definition

Theorem 1 allows us to use our Cournot framework in much the same way as we would use traditional Cournot competition. To complete our treatment and define a game, we must specify what player owns a good at each part of the supply chain. We will say that two graph edges or assembly nodes, $c_1, c_2 \in G \setminus V_M$ are *owner-continuous* if there is a path that includes both that does not include any market nodes in between the elements. This is an equivalence relation, and the resulting equivalence classes denote parts of the supply chain between market nodes, in which a good must be owned by a single player.

Formally, we may define a Cournot supply chain game as an ordered 4-tuple, $(G, \mathbf{t}, Y, \alpha)$, consisting of a supply chain, G, a compatible inverse demand function, $\{t_e(f)\}_{e \in E_M}$, a set of players *Y*, and a ownership map, $\alpha : G \setminus V_M \to Y \cup \{C\}$, such that if graph elements $c_1, c_2 \in G \setminus V_M$ are owner-continuous, $\alpha(c_1) = \alpha(c_2)$, and $\alpha(c) = C$ if and only if the element *c* is owner-continuous with the consumer node. Each player, $y \in Y$, selects a quantity restriction for each market edge $e \in E_M$ with $\alpha(e) = y$.[1] These quantity restrictions determine a unique quantity flow, *f*, and a unique set of market edge prices, $\overline{\mathbf{x}}$, according to Theorem 1. Finally, the payoff to player $y \in Y$ is the sum of the revenues at each market edge she owns, $\sum_{e \in E_M, \alpha(e) = y} f(e) \overline{x}_e$.

4 Network Neutrality Analysis

In the previous section, we presented a formal supply-chain description, and developed a natural way to extend Cournot competition over an entire supply chain. With these tools

[1] Throughout this study, we will assume that players select their quantity restriction simultaneously. The timing may be altered, for example, to denote aggressive players.

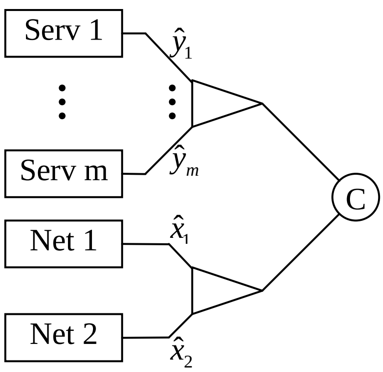
in hand, we now turn our attention to the study of net neutrality. We will examine a series of neutrality regimes, inspired by the literature, depicting each as a supply chain. This will require us to identify the goods exchanged between network actors, and where each is required on the path to a complete final product. We will then apply our Cournot theory to compute equilibria, allowing us to compare prices, welfare, and other metrics.

In order to tie all of our supply chains together, each will share a common set of players, consisting of 2 network providers, NP_1 and NP_2, and *m* service providers, $SP_1,\ldots,SP_m$. Each network provider, NP_i is the owner of a network input, Net *i*. Similarly, each service provider, SP_j is the owner of a service input, Ser *j*. Other inputs may be modeled to represent contractual obligations that occur in different regimes. We will use the variable x_i to represent the quantity of Net *i* and y_j to represent the quantity of Ser *j*. Because a final product will always be built from one network and one service, we will always have $\sum x_i = \sum y_j$. Although our theory admits a highly general demand function, for this analysis we will use a simple linear demand: $t_{ij} = 1 - \sum x_i = 1 - \sum y_j$.

4.1 Zero-price rule

As we described in section 2, most studies employ a zero-price rule as a benchmark representing a neutral network. The key feature common to these models is that service providers do not exchange payment with network providers in order to access customers. Thus, from a supply chain perspective, we need only consider two goods: a network good that is sold from a network provider to the end-user, and a service good that is sold from a service provider to the end-user. We depict these goods as being assembled together by the end-user, which ensures that equal amounts of network and service are produced.

Our supply chain for the zero-price rule is pictured below. In this diagram, inputs are presented at the left, each labeled in a rectangle. Market nodes are depicted as triangles, with a quantity restriction labeled on each incoming edge. The consumer node, in this case, and assembly node, is pictured on the right, labeled by a C. As we described in our theory sections, each firm must choose a quantity restriction instead of a quantity. Let each service provider *j* restrict its input by $\hat{y}_j$, and let each network provider *i* restrict its network input by $\hat{x}_i$.



The total quantity of all final products is $1-\sum \hat{x}_i - \sum \hat{y}_j$, and inverse demand may be written, $t_{ij}(\hat{\mathbf{x}},\hat{\mathbf{y}}) = 1-\sum \hat{x}_{i'} - \sum \hat{y}_{j'}$ for all *i,j*. Since the profits in each market are proportional to the total restriction there, we may compute the price for network *i* to be $\overline{x}_i = \sum \hat{x}_i$, and the price for service *j* to be $\overline{y}_j = \sum \hat{y}_j$. These may also be written in terms of quantities as,

$$\overline{x}_i = 1-\sum x_j - \sum \hat{y}_j,\ \ \overline{y}_i = 1-\sum y_j - \sum \hat{x}_j$$

This relation suggests that the two sides of the market interact by exerting a simple price on each other, exactly what we would expect for firms in a vertical arrangement. In fact, in the case that there is one network provider and one service provider, the game is equivalent to one in which both firms select prices, and the familiar double-marginalization result emerges. Thus, our framework behaves in a very natural way in this scenario.

NP$_i$ makes profit, $\pi_{NP_i} = x_i \overline{x}_i = \left(\frac{1-\sum \hat{y}_j}{2} - \hat{x}_i\right)\sum_k \hat{x}_k$ with first order condition, $\left(\frac{1-\sum \hat{y}_j}{2} - \hat{x}_i\right) - \sum_k \hat{x}_k = 0$. SP*j* makes profit $\pi_{SP_j} = y_j \overline{y}_j = \left(\frac{1-\sum \hat{x}_i}{m} - \hat{y}_j\right)\sum_k \hat{y}_k$ with first order condition, $\left(\frac{1-\sum \hat{x}_i}{m} - \hat{y}_j\right) - \sum_k \hat{y}_k = 0$. These can be solved for the equilibrium,

$$\hat{y}_j = \frac{2}{m(3m+2)},\ \ \hat{x}_i = \frac{m}{(6m+4)} \tag{3}$$

$$x_i = \frac{m}{3m+2},\ \ y_j = \frac{2}{3m+2} \tag{4}$$

$$\overline{x}_i = \frac{m}{3m+2},\ \ \overline{y}_j = \frac{2}{3m+2} \tag{5}$$

$$\pi_{NP_i} = \left(\frac{m}{3m+2}\right)^2,\ \ \pi_{SP_j} = \left(\frac{2}{3m+2}\right)^2 \tag{6}$$

Let p_{zp} be the price of all final products for the end user under the zero-price rule. This is given by,

$$p_{zp} = \overline{x}_i + \overline{y}_i = \frac{m+2}{3m+2} \tag{7}$$

Welfare under this regime, ω_{zp}, may be written,

$$\omega_{zp} = (1-p^2)/2 = \frac{4m(m+1)}{(3m+2)^2} \tag{8}$$

For large *m*, this equilibrium approaches the familiar Cournot equilibrium with 2 firms,

$$x_i = 1/3, \;\; y_j = 0 \tag{9}$$

$$\overline{x}_i = 1/3, \;\; \overline{y}_j = 0 \tag{10}$$

$$\pi_{NP_i} = 1/9, \;\; \pi_{SP_j} = 0 \tag{11}$$

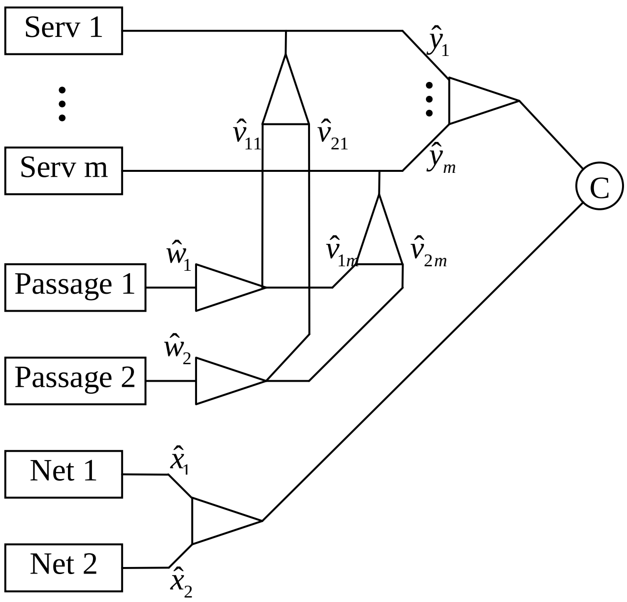
4.2 Uniform Passage Fees

In this section, we consider a widely-discussed scenario, in which each network provider begins directly charging service providers a fee before allowing them to reach its customers. From a supply chain perspective, each ISP creates a right of access to its own customers, which we call passage in order to distinguish it from transit. Unlike transit, a service provider must purchase passage from every network provider with customers it transmits to.

When incorporating passage into our Cournot framework, a key challenge is ensuring that each service provider purchases the right amount of passage from each network provider, corresponding to the number of end-users connected to that provider. Our model provides a convenient way to accomplish this: Each service provider can choose the right amount of each passage good in an *internal market*. In effect, we will allow each service provider to sell passage to itself, meaning that no money will be exchanged between players. The restrictions in this internal market will be set to ensure the proper amount of each passage good, and the total restriction will be zero, ensuring that this market does not affect the service provider's profits.

Our strategy is shown in more detail in the supply chain below. As passage goods move along this supply chain, they are first sold from a network provider to a service provider. They next enter an internal market, in which each service provider *i* selects between passage goods, adding quantity restrictions $\{\hat{v}_{1i}, \hat{v}_{2i}\}$, which we will compute to ensure the proper amount of each good, with $\hat{v}_{1i} + \hat{v}_{2i} = 0$. We will assume that the price of a final product does not depend on which passage input it includes. This will ensure that the total price of each passage input is the same when it is combined with a service input, and each service provider therefore charges a single price at the end-user market.

If our methodology strikes the reader as unorthodox, note that when we restrict to our linear demand function, our model is fully equivalent to one in which network providers select prices for passage and charge these to service providers. It would be possible to use prices in this way for a general demand function, but our use of internal markets allows us to maintain a consistent treatment of markets as Cournot.



Claim 1: The equilibrium solution for the Uniform Passage Fees supply chain is as follows. Prices are given by,

$$\hat{x} = -\frac{m}{2(3m+1)}, \quad \hat{w} = \frac{3m}{2(3m+1)}, \quad \hat{y} = \frac{1}{m(3m+1)} \tag{12}$$

With quantities,

$$x_i = \frac{m}{2(3m+1)}, \quad y_i = \frac{1}{3m+1}, \quad v_{ij} = \frac{1}{2(3m+1)}, \quad w_i = \frac{m}{2(3m+1)} \tag{13}$$

And prices,

$$\bar{x}_i = -\frac{m}{3m+1}, \quad \bar{y}_i = \frac{1}{3m+1}, \quad \bar{w}_i = \frac{3m}{3m+1}, \quad \bar{v}_{ij} = 0 \tag{14}$$

A proof is given in the appendix. The price for end users in the uniform passage fees regime, p_{upf} , is given by,

$$p_{upf} = \bar{x} + \bar{y} + \bar{v} + \bar{w} = \frac{2m+1}{3m+1} \tag{15}$$

which is considerably higher than the price under the zero-price rule, $\frac{m+2}{3m+2}$. In fact, in the limit that the service industry is fully competitive, $m \to \infty$, the price for end-users is twice as high under uniform passage fees.

Moreover, the price increase compared with the zero-price rule, $\frac{3m^2}{(3m+1)(3m+2)}$, increases with *m*, meaning that the extra price from the move away from neutrality is greatest when the service market is competitive.

It is also interesting to note that the price for network service in the end-user market is actually negative – network actually pay end-users to select their networks. Intuitively, it is better for a network provider to earn revenue by charging service providers, knowing that this lowers the demand for both networks easily. At the same time, the provider pays end-users hoping to establish as much market share as possible.

A similar effect occurs in the familiar credit card market. Credit companies set prices for card-holders to zero, even offering incentives to make the price effectively negative, preferring to charge positive fees from merchants and other sources.

Equilibrium profits are given by,

$$\pi_{NP_i} = \frac{m^2}{(3m+1)^2}, \quad \pi_{SP_i} = \frac{1}{(3m+1)^2} \tag{16}$$

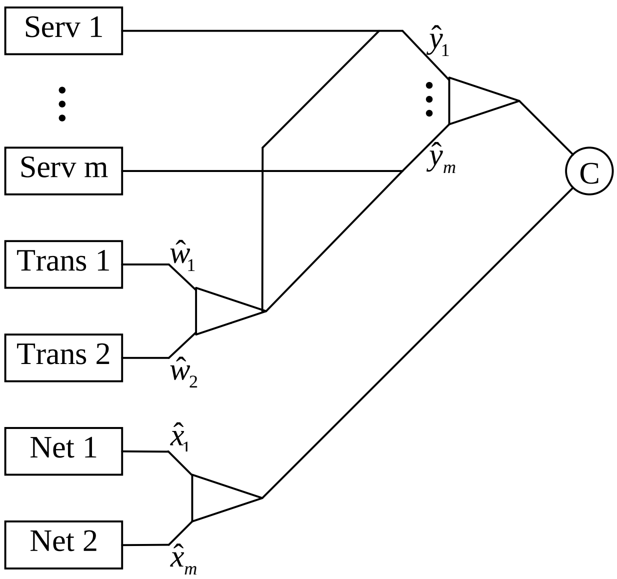
Welfare in this regime, ω_{upf}, may be written,

$$\omega_{upf} = (1 - p_{dp}{}^2)/2 = \frac{5m^2 + 2m}{2(3m+1)^2} \tag{17}$$

This is considerably lower than welfare under the zero-price rule. In fact, in the limiting case of a fully competitive service industry, welfare is only 5/8 as high.

4.3 Duopoly price rule

To some extent, the neutrality debate concerns the network providers' desire to extract money from service providers. Passage fees are a widely-discussed method for network providers to accomplish this, but they are not the only method. In this section, we consider what would happen if the network providers rewrote their peering agreements with other ISPs, and began to charge them money for interconnection. This would result in duopoly competition, in which each service provider would need to buy transit from one of the two network providers. This scenario is pictured in the supply chain below.



Claim 2: The equilibrium for the Duopoly-price rule supply chain is as follows: Equilibrium quantities are,

$$x_i = \frac{3m}{2(5m+3)}, \quad y_i = \frac{3}{5m+3}, \quad w_i = \frac{3m}{2(5m+3)} \tag{18}$$

With equilibrium prices,

$$\overline{x}_i = \frac{m}{5m+3}, \quad \overline{y}_i = \frac{3}{5m+3}, \quad \overline{w}_i = \frac{m}{5m+3} \tag{19}$$

A proof is given in the appendix. Let p_{dp} be the price of all final products for the end user under the duopoly-price rule. This is given by,

$$p_{dp} = \overline{x}_i + \overline{y}_i + \overline{w}_i = \frac{2m+3}{5m+3} \tag{20}$$

which is strictly larger than the end-user price under the zero-price rule, but smaller than the end-user price under uniform passage fees. As before, the extra price compared to the zero-price rule, $\frac{2m+3}{5m+3} - \frac{m+2}{3m+2} = \frac{m^2}{(5m+3)(3m+2)}$, increases with *m*. In other words, the increase in price from moving to the less neutral regime is most pronounced when the service industry is competitive.

Equilibrium profits are,

$$\pi_{NP_i} = \frac{3m^2}{(5m+3)^2}, \quad \pi_{SP_i} = \frac{9}{(5m+3)^2} \tag{21}$$

Welfare in this regime, ω_{dp}, may be written,

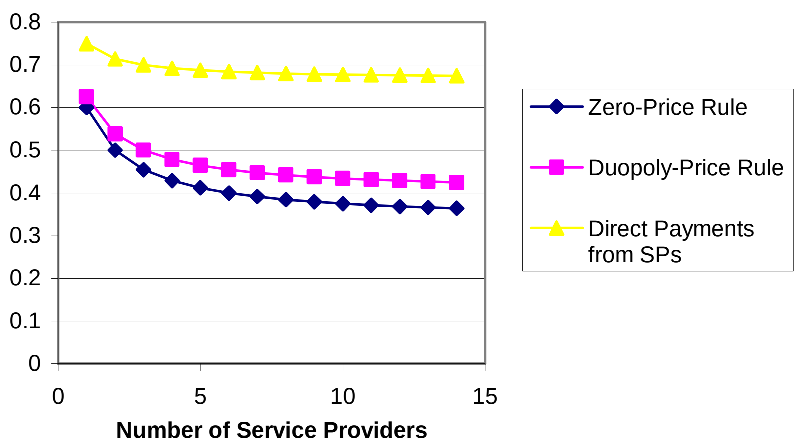
$$\omega_{dp} = (1-p^2)/2 = \frac{3m(7m+6)}{2(5m+3)^2} \tag{22}$$

The loss in welfare compared to the zero-price rule, $\omega_{dp} - \omega_{zp}$, similarly increases with *m*. This means that the welfare loss from non-neutrality is most pronounced when the service industry is competitive.

4.4 Comparison of Neutrality Regimes

In the first graph below, we compare the total end-user price across all three regimes we analyze. Much of the neutrality debate centers around the possibility that network providers introduce passage fees. As the graph shows, the end-user price under this regime is considerably higher than the two alternatives we tested. This is due to the double-marginalization effect we identified previously. Prices are also elevated under the duopoly-price rule, but by a far smaller amount. In this scenario, network providers compete in a partially vertical arrangement. Restricting quantity in the network market also reduces a competitor's output in the transit market, and vice-versa, partially aligning incentives.

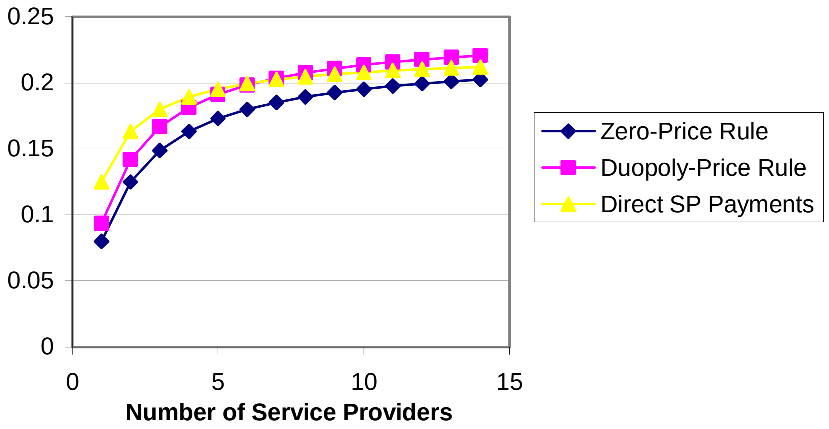
Figure 2: End-user prices under three neutrality regimes



In the next figure, we plot network provider profits under our three neutrality regimes. As one might expect from the elevated prices, network providers earn higher profits under both uniform passage fees and the duopoly price rule. It is interesting to note that the scenario preferred by the network providers depends on the number of service providers. When the number of service providers is small, they are able to extract higher rents, squeezing the potential profit margins of the network providers. Under uniform passage fees, however, double-marginalization by network providers prevents service providers from charging high prices, somewhat counteracting this effect. For this reason, uniform passage fees give network providers a particular advantage when the service market is concentrated.

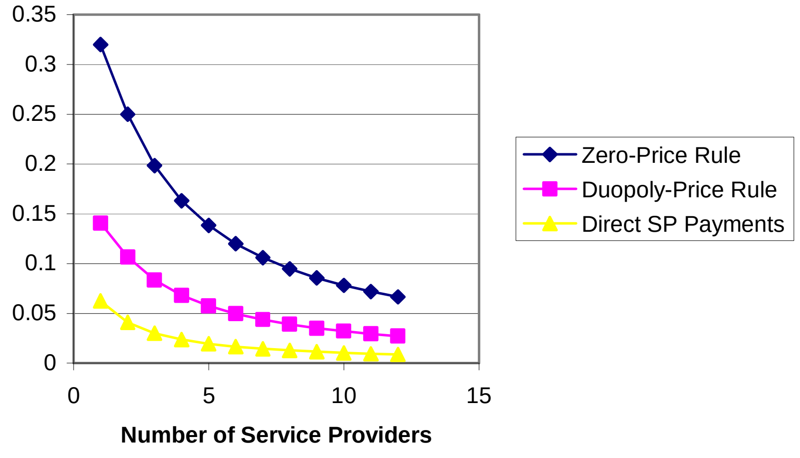
Under a competitive service market, on the other hand, service prices drop, and the duopoly price rule gives network providers an advantage. Their partially vertical arrangement aligns incentives, and improves their long-run profits.

Figure 3: Network provider profits under three neutrality regimes



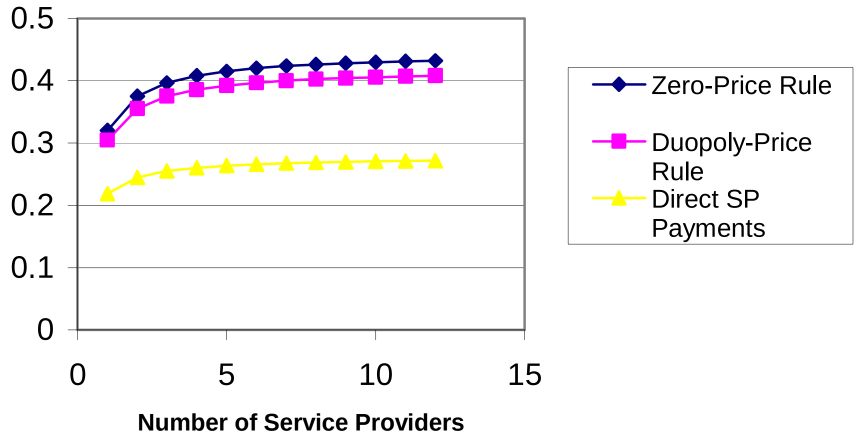
The next graph depicts service provider profits in our regimes. As expected, profits fall as the service industry becomes more competitive. Furthermore, there is a clear and dramatic difference in service profits between the three regimes, with the zero-price rule providing the most profits, and uniform passage fees the least.

Figure 4: Service provider profits under three neutrality regimes



Finally, we may compare welfare between our regimes in the graph below. The zero-price rule features the highest welfare. By comparison, uniform passage fees result in a dramatic drop in welfare, owing to the double-marginalization effect. The duopoly price rule, however, appears as an attractive alternative, producing nearly as much welfare as the zero-price rule.

Figure 5: Total welfare under three neutrality regimes



5 Discussion

At first glance, our supply chain framework may appear to be a complex way to arrive at net neutrality results. Certainly, a lot of machinery goes into developing our Cournot theory. The rewards, however, are considerable. We are able to treat a great variety of regimes under a uniform framework, with tractable expressions, potential for further expressive assumptions, and a natural treatment of market power. Moreover, our framework is technology-agnostic, guaranteeing that results are driven exclusively by market structure, not specific assumptions about technologies. Through these efforts, we are able to start building a more comprehensive picture of the larger neutrality space.

Moreover, our study's potential for broader impact is tremendous. Supply chains exist in countless other industries, and our framework provides a general-purpose tool that can be put to use in many scientific analyses. By using a demand function calibrated with empirical measurements, managers may apply our framework to predict behavior under realistic market conditions. Regulators too, may find potential in our theory, understanding the effects of mergers in supply-chain industries.

Our study of net neutrality has provided fertile ground in which to test our Cournot theory. Much of the neutrality debate has been driven by a seemingly straightforward question: what would the effects be of abandoning net neutrality in the internet. By comparing a larger number of regimes in a unified framework, our study suggests that the answer depends on exactly what type of regime replaces neutrality.

As we have seen, uniform passage fees create a double-marginalization effect that drive up prices and considerably reduce welfare. At the same time, our analysis points towards a duopoly-price rule as an attractive alternative. Such a regime produces nearly as much welfare as the zero-price rule, which network provider extract nearly as much, or sometimes even more, profit as uniform passage fees allow.

As we continue to pursue our research, we will explore still more examples of neutrality regimes, and also add new metric to use in their comparison. In particular, we will soon begin to study incentives for firms to invest in service innovations and network upgrades under each regime. Through these efforts, we aim to build a fuller picture of the larger neutrality space.

6 References

Cheng, H., Bandyopadhyay, S., & Guo, H. The Debate on Net Neutrality: A Policy Perspective. *Information Systems Research*. Vol. 22, No. 1, March 2011, pp. 60-82.

Choi, J. P., & Kim, B. 2010. Net neutrality and investment incentives. *RAND Journal of Economics*, RAND Corporation, vol. 41(3), pp. 446-471.

Economides, N. & Tåg, J. Net Neutrality on the Internet: A Two-Sided Market Analysis (May 2009). NET Institute Working Paper No. 07-45. Available at SSRN: http://ssrn.com/abstract=1019121

Hermalin, B., & Katz, M. L. (2007). The economics of product-line restrictions with an application to the network neutrality debate. *Information Economics and Policy*, *19*(2), pp. 215-248.

Musacchio, J., Schwartz, G., & Walrand, J, A Two-sided Market Analysis of Provider Investment Incentives with an Application to the Net-Neutrality Issue. *Review of Network Economics*. vol. 8, no. 1, March 2009, pp. 22-39.

Norton, W. Internet Transit Prices - Historical and Projected. Dr. Peering white paper. www.drpeering.net. Aug 2010.

7 Appendix

Proof of Lemma 1. Suppose that we are given values for a flow f at C and at all incoming edges except one for each market node. We may then induct backwards to compute the flow at any component, c. If c is a node, and we know the flow at its outgoing edges, the flow at c must be the sum of these. If c is a non-market edge, and we know the flow at its child node, the flow at c must be the same value. If c is a market-edge, and we know the flow at its child node, we are either given the flow at c by assumption, or we may uniquely compute it since the flow at c's child node must be the sum of the flows at its incoming edges. Finally, we have specified $|E_M|-|V_M|+1$ values by assumption, and any choice of these values yields a different flow, so this must be the dimension of *F*. □

Proof of Theorem 1. Let $F/f=\{f'\in F, f'\perp f\}$ be the orthogonal complement of *f* in *F*. Choose a basis, $\{g_i\}$ for F/f. For each market node, $m\in V_M$, let $f_m\in \mathrm{E}_M$ be the vector that's equal to *f* at *m*'s incoming edges, and zero everywhere else. As before, let $\theta: F\to \mathrm{E}_M$, $\theta(f)(e)=f(e)$, be the orthogonal projection taking flows into E_M.

Lemma 4: $\{\theta(g_i)\}\cup\{f_m\}$ is a basis for E_M.

Proof. We first show that $\{\theta(g_i)\}\cup\{f_m\}$ are linearly independent. The $\{g_i\}$ are linearly independent, and θ is a one-to-one linear transformation so the $\{\theta(g_i)\}$ are linearly independent. The $\{f_m\}$ are also linearly independent, so it is sufficient to show that no linear combination of the $\{f_m\}$ yields a non-zero linear combination of the $\{\theta(g_i)\}$, or equivalently, $\theta(f')$ for some $f'\in F/f$. Assume for contradiction that there is such an $f'\in F/f$, with $\theta(f')=\sum_{m\in V_M} a_m f_m$ for some constants, $\{a_m\}$. Without loss of generality, we may scale *f'* so that $f'(C)=f(C)$. Assume for induction that $f'(c)=f(c)$ for some graph component *c*. If *c* is an edge, and *n* is its parent node, to be valid, the flow at *n* must be the flow at *c* times the number of *n*'s outgoing links, so we have $f'(n)=f(n)$. If *c* is an assembly node, the flow at any incoming edge, $e\in\mathrm{inc}(c)$ must be the flow at *c*, so we have $f'(e)=f(e)$. Finally, if *c* is a market node, we know that $f'(e)=\theta(f')(e)=\sum a_m f_m(e)=a_c f_c(e)$ for all $e\in\mathrm{inc}(c)$. But for f' to be a valid flow, we must have $\sum_{e\in\mathrm{inc}(c)} f'(e)=a_c\sum_{e\in\mathrm{inc}(c)} f_c(e)=f'(c)=f(c)$. Since *f* is a valid flow, $\sum_{e\in\mathrm{inc}(c)} f_c(e)=f(c)$ and therefore $a_c=1$. Hence, $f'(e)=f(e)$ for all $e\in\mathrm{inc}(c)$. We may conclude that $f'=f$, contradicting $f'\in F/f$.
Finally, since *F* has dimension $|E_M|-|V_M|+1$, F/f has dimension $|E_M|-|V_M|$. Thus, this is also the number of the $\{\theta(g_i)\}$, while there are $|V_M|$ of the $\{f_m\}$, for a total of $|E_M|$ linearly independent vectors. This is also the dimension of E_M, so $\{\theta(g_i)\}\cup\{f_m\}$ is a basis for E_M. □

First, assume that the price function $\bar{\mathbf{x}}$ fulfills assumption 1. Since the price of each final product is the same under price vectors $\bar{\mathbf{x}}$ and $\mathbf{t}$, and the $\{u_d\}$ span *F*, the total price of any flow must be the same as well.

$$\bar{\mathbf{x}}\cdot f'=\mathbf{t}\cdot f' \text{ for all } f'\in F \tag{23}$$

Since the vectors $\{\theta(g_i)\} \cup \{f_m\}$ form a basis for E_M, $\bar{\mathbf{x}}$ is uniquely identified by its dot products with these vectors. Furthermore, $\bar{\mathbf{x}}$ may be written as a linear function of these dot products. Since the g_i lie in *F*, we have $\bar{\mathbf{x}} \cdot \theta(g_i) = \bar{\mathbf{x}} \cdot g_i = \mathbf{t} \cdot g_i$ for all *i*. $\bar{\mathbf{x}} \cdot f_m$ is simply the revenue at market node *m*, which we label π_m. Moreover, we have $\sum \pi_m = \sum \bar{\mathbf{x}} \cdot f_m = \bar{\mathbf{x}} \cdot \sum f_m = \bar{\mathbf{x}} \cdot f$. As long as this equals $\mathbf{t} \cdot f$, equation (23) will hold for *f* as well as $\{g_i\}$, and therefore all of F. Thus, any set of revenues, $\{\pi_m\}$, with $\sum \pi_m = \mathbf{t} \cdot f$ identifies a unique $\bar{\mathbf{x}}$ fulfilling demand over F. We therefore need only show that there exists a unique set of market node profits, $\{\pi_m\}$, with $\sum \pi_m = \mathbf{t} \cdot f$, such that the resulting price function $\bar{\mathbf{x}}$ fulfills assumptions 2,3,and 4.

Next, assume a function $\{\pi_m\}$ with $\sum \pi_m = \mathbf{t} \cdot f$, such that the resulting price function fulfills assumption 2. Fix a flow, *f*. By Lemma 3, for each market node, *m*, we can uniquely choose a restriction vectors, $\hat{\mathbf{u}}_m$, such that $\sum_{e \in \mathrm{inc}(m')} \hat{\mathbf{u}}_m(e) = 0$ for all $m' \neq m$, and $\omega(\hat{\mathbf{u}}_m) = f$. For a general restriction that induces *f*, $\hat{\mathbf{x}} \in \omega^{-1}(f)$, let $\alpha_m = \dfrac{\sum_{e \in \mathrm{inc}(m)} \hat{x}_e}{\sum_{e \in E_M} \hat{x}_e}$ be the fraction of the total restriction located at market node *m*. Since $\omega^{-1}(f)$ is an affine space, the affine combination $\sum_{m \in V_M} \alpha_m \hat{\mathbf{u}}_m$ is also in $\omega^{-1}(f)$. Furthermore, the total restriction at a market node *m* for this vector is

$$\alpha_m \sum_{e \in \mathrm{inc}(m)} \hat{\mathbf{u}}_m(e) = \frac{\sum_{e \in \mathrm{inc}(m)} \hat{x}_e}{\sum_{e \in E_M} \hat{x}_e} \sum_{e \in E_M} \hat{\mathbf{u}}_m(e) = \frac{\sum_{e \in \mathrm{inc}(m)} \hat{x}_e}{f(C)} f(C) = \sum_{e \in \mathrm{inc}(m)} \hat{x}_e,$$

which is the same as for $\hat{\mathbf{x}}$. By Lemma 2, there is only one vector with this property, so $\hat{\mathbf{x}} = \sum_{m \in V_M} \alpha_m \hat{\mathbf{u}}_m$.

Since prices, and therefore profits, are linear within this domain, we may write the relation,

$$\pi_m(\hat{\mathbf{x}}) = \sum_{m' \in V_M} \alpha_{m'} \pi_m(\hat{\mathbf{u}}_{m'}) . \tag{24}$$

Thus, π_m is completely determined by the values, $\{\pi_m(\hat{\mathbf{u}}_{m'})\}$. Conversely, given any set of values, $\{\pi_m(\hat{\mathbf{u}}_{m'})\}$, with $\sum_{m \in V_M} \pi_m(\hat{\mathbf{u}}_{m'}) = \mathbf{t} \cdot f$ for all *m'*, we may extend it linearly in the manner of (24) to create a profit functions $\{\pi_m\}$, with $\sum \pi_m = \mathbf{t} \cdot f$. $\bar{\mathbf{x}}$ is linear in the $\{\pi_m\}$, which are linear in $\hat{\mathbf{x}}$ for fixed *f*, so the price function determined in this way fulfills assumption 2. We therefore need only show that there exist unique values for

$\{\pi_m(\hat{\mathbf{u}}_{m'})\}$, with $\sum_{m\in V_M} \pi_m(\hat{\mathbf{u}}_{m'}) = \mathbf{t}\cdot f$, such that the resulting price function fulfills assumption 3 and 4.

Next, assume a set of values, $\{\pi_m(\hat{\mathbf{u}}_{m'})\}$, with $\sum_{m\in V_M} \pi_m(\hat{\mathbf{u}}_{m'}) = \mathbf{t}\cdot f$, for every flow *f*, fulfilling assumptions 3 and 4. Given flow *f*, and corresponding $\hat{\mathbf{u}}_{m'}$, we consider changing the restrictions at a single node $m \neq m'$, leaving the total restriction constant. This is equivalent to adding a vector from $B_m = \{\hat{\mathbf{y}} : \hat{y}_e = 0, e \notin \text{inc}(m), \sum \hat{y}_c = 0\}$. Let $\hat{\mathbf{y}}_m \in B_m$ be a vector that maximizes total profit when added to $\hat{\mathbf{u}}_{m'}$,

$$\hat{\mathbf{y}}_m = \arg\max_{\hat{\mathbf{y}}} \pi(\hat{\mathbf{u}}_{m'} + \hat{\mathbf{y}}) \tag{25}$$

Since ω is an affine transformation, $\omega(\hat{\mathbf{x}}+\hat{\mathbf{y}})$ is the same flow for any $\hat{\mathbf{x}}$ with $\omega(\hat{\mathbf{x}}) = f$. Total profit is a function of flow, so we can replace $\hat{\mathbf{u}}_{m'}$ with any such $\hat{\mathbf{x}}$ without affecting our maximization problem. In particular, for any choice of m', $\hat{\mathbf{y}}_m$ maximizes $\pi(\hat{\mathbf{u}}_{m'} + \hat{\mathbf{y}})$. Let $\hat{\mathbf{u}}_{m'}|_{B_m} = \hat{\mathbf{u}}_{m'} + y_m$ be the resulting restriction vector for each $\hat{\mathbf{u}}_{m'}$. Let $\delta_m(f)$ be the extra total profit from changing the restrictions at *m* in this manner.

$$\delta_m(f) = \pi(\hat{\mathbf{u}}_{m'}|_{B_m}) - \pi(\mathbf{u}_{m'}) = \pi(\hat{\mathbf{x}} + \hat{\mathbf{y}}_m) - \pi(\hat{\mathbf{x}}), \text{ any } \hat{\mathbf{x}} \in \omega^{-1}(f) \tag{26}$$

Loosely speaking, we can think of this quantity as the loss in profit due to suboptimal selection of goods at market *m*.

Assumption 3 tells us that changing the restrictions at node *m* while keeping the total restriction constant does not affect the sum of the profits at the other market nodes,

$$\sum_{n\neq m} \pi_n(\hat{\mathbf{u}}_{m'}|_{B_m}) = \sum_{n\neq m} \pi_n(\hat{\mathbf{u}}_{m'}) \tag{27}$$

Expanding the total profit, we may write,

$$\pi(\hat{\mathbf{u}}_{m'}|_{B_m}) - \pi(\hat{\mathbf{u}}_{m'}) = \sum_{n\in V_M} \pi_n(\hat{\mathbf{u}}_{m'}|_{B_m}) - \pi_n(\hat{\mathbf{u}}_{m'}) = \pi_m(\hat{\mathbf{u}}_{m'}|_{B_m}) - \pi_m(\hat{\mathbf{u}}_{m'}) \tag{28}$$

Since $\hat{\mathbf{u}}_{m'}|_{B_m}$ maximizes the left-hand side over restrictions at *m* that maintain zero total restriction, it also maximizes π_m. By assumption 4, we know $\pi_m(\hat{\mathbf{u}}_{m'}|_{B_m}) = 0$, so we can write,

$$\pi_m(\hat{\mathbf{u}}_{m'}) = \pi(\hat{\mathbf{u}}_{m'}) - \pi(\hat{\mathbf{u}}_{m'}|_{B_m}) = -\delta_m \tag{29}$$

As for the profit at *m'*, we may write,

$$\pi_{m'}(\hat{\mathbf{u}}_{m'}) = \pi(\hat{\mathbf{u}}_{m'}) - \sum_{m\neq m'} \pi_m(\hat{\mathbf{u}}_{m'}) = \pi(\hat{\mathbf{u}}_{m'}) + \sum_{m\neq m'} \delta_m \tag{30}$$

Extending according to (24), we may write the general profit function,

$$\begin{aligned}\pi_m(\hat{\mathbf{x}}) &= \sum_{m'\in V_M} \alpha_{m'}\pi_m(\hat{\mathbf{u}}_{m'}) \\ &= \alpha_m\left[\pi(\hat{\mathbf{u}}_m) + \sum_{m'\neq m}\delta_{m'}\right] - \sum_{m'\neq m}\alpha_{m'}\delta_m \, . \\ &= \alpha_m\left[\pi(f) + \sum_{V_M}\delta_{m'}\right] - \delta_m\end{aligned} \tag{31}$$

Roughly speaking, $\pi(f)+\sum_{V_M}\delta_{m'}$ is the excess profit after removing the profit loss from suboptimal allocation at each market. It is this excess that is allocated proportionally to each market according to the total restrictions.

Conversely, given a set of profit functions defined according to (31), for a restriction $\hat{\mathbf{x}}$, with $\sum_{e\in \mathrm{inc}(m)} \hat{x}_e = 0$ for some node *m*, we have $\pi_m(\hat{\mathbf{x}}) = -\delta_m\left(\omega(\hat{\mathbf{x}})\right)$. Given a vector $\hat{\mathbf{y}} \in B_m$, we also have, $\pi_m(\hat{\mathbf{x}}+\hat{\mathbf{y}}) = -\delta_m\left(\omega(\hat{\mathbf{x}}+\hat{\mathbf{y}})\right)$. For this new vector, the total revenue at all markets other than *m*, can be written,

$$\begin{aligned}&\pi(\hat{\mathbf{x}}+\hat{\mathbf{y}}) - \pi_m(\hat{\mathbf{x}}+\hat{\mathbf{y}}) = \pi(\hat{\mathbf{x}}+\hat{\mathbf{y}}) + \delta_m\left(\omega(\hat{\mathbf{x}}+\hat{\mathbf{y}})\right) \\ &= \pi(\hat{\mathbf{x}}+\hat{\mathbf{y}}) + \max\nolimits_{\hat{\mathbf{z}}\in B_m} \pi(\hat{\mathbf{x}}+\hat{\mathbf{y}}+\hat{\mathbf{z}}) - \pi(\hat{\mathbf{x}}+\hat{\mathbf{y}}) \\ &= \max\nolimits_{\hat{\mathbf{z}}\in B_m} \pi(\hat{\mathbf{x}}+\hat{\mathbf{z}}) = \delta_m\left(\omega(\hat{\mathbf{x}})\right) + \pi(\hat{\mathbf{x}}) = \pi(\hat{\mathbf{x}}) - \pi_m(\hat{\mathbf{x}})\end{aligned} \tag{32}$$

Which is the total revenue at all other markets under the original vector, $\hat{\mathbf{x}}$, fulfilling assumption 3. Furthermore, we know $\delta_m(\omega(\hat{\mathbf{x}})) = \max_{\hat{\mathbf{y}}\in B_m} \pi(\hat{\mathbf{x}}+\hat{\mathbf{y}}) - \pi(\hat{\mathbf{x}})$. Letting $\hat{\mathbf{y}}_m \in B_m$ be a vector that maximizes this quantity, we then have $\delta_m(\omega(\hat{\mathbf{x}}+\hat{\mathbf{y}}_m)) = \max_{\hat{\mathbf{y}}\in B_m} \pi(\hat{\mathbf{x}}+\hat{\mathbf{y}}_m+\hat{\mathbf{y}}) - \pi(\hat{\mathbf{x}}+\hat{\mathbf{y}}_m) = 0$, since $\hat{\mathbf{y}}+\hat{\mathbf{y}}_m \in B_m$ and $\hat{\mathbf{y}}_m$ already maximizes profit. Then $\pi(\hat{\mathbf{x}}+\hat{\mathbf{y}}_m) = -\delta_m(\omega(\hat{\mathbf{x}}+\hat{\mathbf{y}}_m)) = 0$. Moreover, δ_m is defined to be non-negative, so this is the maximum possible revenue at *m*, fulfilling assumption 4. We have therefore identified a unique set of values, $\{\pi_m(\hat{\mathbf{u}}_{m'})\}$, with $\sum_{m\in V_M} \pi_m(\hat{\mathbf{u}}_{m'}) = \mathbf{t}\cdot f$, fulfilling assumptions 3 and 4, which completes the proof. □

Proof of Lemma 2. It is easy to see that ϕ is a linear transformation between vector spaces. Let $H\,/\,Ker\ \phi$ be its coimage. ϕ is one-to-one over its coimage, so we may write $\phi^{-1}: F \to H\,/\,Ker\ \phi$ for its right-inverse. We may then define $\psi: F \to \mathbb{R}^{|D|}, \psi(f) = \mathbf{r}(\phi^{-1}(f))$. For any $h \in H$, we have $\phi\left(\phi^{-1}\circ\phi(h)\right) = \phi(h)$, so assumption (a) implies that prices are the same for quantities $\phi^{-1}\circ\phi(h)$ and *h*,

$$\mathbf{r}(h) = \mathbf{r}(\phi^{-1}\circ\phi(h)) = \psi\circ\phi(h) \tag{33}$$

Given a set of prices, $\mathbf{x} \in \mathbb{R}^{|D|}$, let $\pi_{\mathbf{x}} : H \to \mathbb{R}^1$, $\pi_{\mathbf{x}}(h) = \mathbf{x}^T h$ be the linear transformation that takes quantities to total revenue. For any $c \in Ker\ \phi$, assumption (b) yields

$$\pi_{\mathbf{r}(h)}(h) = \pi_{\mathbf{r}(h+c)}(h+c) = \pi_{\mathbf{r}(h)}(h+c) = \pi_{\mathbf{r}(h)}(h) + \pi_{\mathbf{r}(h)}(c) \tag{34}$$

Which implies that $\pi_{\mathbf{r}(h)}(c) = 0$. For quantity vector $i \in H$, $i = \phi^{-1} \circ \phi(i) + c_i$ for some $c_i \in Ker\ \phi$, so we may write,

$$\pi_{\mathbf{r}(h)}(i) = \pi_{\mathbf{r}(h)}(\phi^{-1} \circ \phi(i) + c_i) = \pi_{\mathbf{r}(h)}\left(\phi^{-1} \circ \phi(i)\right) + \pi_{\mathbf{r}(h)}(c_i) = \pi_{\mathbf{r}(h)} \circ \phi^{-1} \circ \phi(i) \tag{35}$$

We need to describe the space of flows in terms of the flow at each market edge. To do this, let $\mathrm{E}_M = \{\varepsilon : E_M \to \mathbb{R}\}$ be the vector space of arbitrary real numbers assigned to the market edges. We may define linear transformation $\theta : F \to \mathrm{E}_M$, $\theta(f)(e) = f(e)$, which takes flows into this larger vector space. It is easy to see that θ is one-to-one, so we may choose a left-inverse, $\theta^{-1} : \mathrm{E}_M \to F$, such that $\theta^{-1} \circ \theta(f) = f$. Substituting into the above,

$$\pi_{\mathbf{r}(h)}(i) = \pi_{\mathbf{r}(h)} \circ \phi^{-1} \circ \theta^{-1} \circ \theta \circ \phi(i) \tag{36}$$

Letting $\rho_f = \pi_{\psi(f)} \circ \phi^{-1} \circ \theta^{-1}$, this simplifies to,

$$\pi_{\mathbf{r}(h)}(i) = \rho_{\phi(h)}\left(\theta \circ \phi(i)\right) \tag{37}$$

Let $v_e \in \mathrm{E}_M$ be the unit vector which assigns 1 to edge *e* and zero to all other edges. The set $\{v_e\}_{e \in E_M}$ is a basis for E_M, so we may express $\rho_{\phi(h)}$ in this basis as some matrix,

$$T_h = [t_1, ..., t_{|E_M|}] \tag{38}$$

For some market-edge prices, $\{t_e : F \to \mathbb{R}\}_{e \in E_M}$. Revenue can then be expressed as,

$$\pi_{\mathbf{r}(h)}(i) = [t_1, ..., t_{|E_M|}]\left[\theta \circ \phi(i)(e)\right]_{e \in \mathrm{E}_M} = \sum_{e \in E_M} t_e\left(\theta \circ \phi(i)(e)\right) = \sum_{e \in E_M} t_e\left(\phi(i)(e)\right) \tag{39}$$

Finally, the price of final product *d* can be computed as the revenue for vector $i_d \in H$, which is one for final product *d* and zero for all other products,

$$\pi_{\mathbf{r}(h)}(i_d) = \sum_{e \in E_M} t_e\left(\phi(i_d)(e)\right) = \sum_{e \in d \cap E_M} t_e\left(\phi(h)\right), \tag{40}$$

which is the required form. □

Proof of Lemma 3. First, we show that the set of total restrictions, $\{\hat{s}_m\}$, uniquely determines the upstream restriction at every component that is not a market edge. The upstream restrictions of the input nodes are uniquely determined (to be zero). We proceed inductively to show that the upstream restriction at a component, *c*, is uniquely determined. There are three cases to consider:

1. *c* is an assembly node. If the upstream restriction at *c*'s incoming is uniquely determined, the upstream restriction at *c* is simply their sum, so it is uniquely determined.
2. *c* is a non-market edge. If the upstream restriction at *c*'s parent node is uniquely determined, the upstream restriction at *c* is the parent's upstream restriction divided by the number of its outgoing edges, so it is uniquely determined.

3. c is a market node. If the upstream restriction is uniquely determined at the parent nodes of c's incoming edges, the upstream restriction at an incoming edge, $e \in \text{inc}(c)$ is found by adding a fraction of the upstream restriction at e's parent to the restriction at e. The upstream restriction at c is the sum of all these upstream restrictions, which is the sum of these fractional components (which are uniquely determined) and the restrictions at each of c's incoming edges, which is simply $\hat{s}_c$. Thus, c's upstream restriction is uniquely determined.

Next we show that there is a unique upstream restriction vector, $\hat{\mathbf{w}}$, compatible with total market restrictions $\{\hat{s}_m\}$ which results in flow f. By the first step, we know that $\hat{\mathbf{w}}$ is uniquely determined at each node that is not a market node. In order for $\hat{\mathbf{w}}$ to result in flow f, we must have $f(C) = N - \hat{w}_C = N - \sum \hat{s}_m$, which is true by assumption, and for each market node, there must be a constant c_m such that if e is an incoming edge of m, $f(e) = c_m - \hat{w}_e$. We must also have $\sum_{e \in \text{inc}(m)} \hat{w}_e = \hat{w}_m$ for $\hat{\mathbf{w}}$ to be a value upstream restriction vector, so $c_m = \dfrac{f(m) + \hat{w}_m}{|\text{inc}(m)|}$ and $\hat{w}_e = \dfrac{f(m) + \hat{w}_m}{|\text{inc}(m)|} - f(e)$. In this way, we may uniquely determine all the $\{\hat{w}_c\}$. Finally, for each market edge, e, we may compute $\hat{x}_e$ as the difference between the upstream restriction at e's parent divided by the number of its outgoing edges, and the upstream restriction at e. Thus, we have uniquely determined the restriction vector, $\hat{\mathbf{x}}$, with market totals $\{\hat{s}_m\}$ which yields flow f. □