EIA Global Hydrocarbon Supply Model—Transportation and Logistics Modeling

Component Design Report (CDR)

By

Dr. Dale M. Nesbitt
dale.nesbitt@arrowheadeconomics.com
dnesbitt@stanford.edu

ArrowHead Economics

27121 Adonna Ct.
Los Altos Hills, CA 94022
(650) 218-3069 mobile

www.arrowheadeconomics.com

November 1, 2014

ACKNOWLEDGMENTS AND BIOGRAPHY

The author is grateful for the reviews and analysis by Dr. Antoine Calvez, Robert Stibolt, Dr. Jill Scotcher, Dr. Hill Huntington, Dr. John Weyant, and Randy Begotka. They have helped to elucidate key points and spot errors of commission or omission, bring valuable analytical and modeling perspective, and ensure that the mathematics and concepts are correct.

Biography

Dr. Nesbitt studied operations research, probability, and economics in Management Science and Engineering at Stanford. His dissertation (**Policy Ordering in semi-Markov Decision Processes** in Engineering-Economic Systems at Stanford University) contained a unique application of linear programming and probability in the semi-Markov arena. (It was unique in the sense that linear programming was used as an analytical method of proof as well as a solution method.) He has applied nonlinear programming and Kuhn-Tucker conditions to a wide range of optimization problems inside and outside the economic realm. He has applied linear and nonlinear programming to a wide array of problems ranging from nonlinear maximum likelihood estimation (in the probability and consumer choice space) to intertemporal storage (in the economic space) to pricing and revenue management (in the economic space) to producer profit maximization/consumer utility maximization (in the economic space) to Koopmans-Hitchcock activity analysis (in the fleet management space). His optimization and operations research credentials, often published, have created experience and understanding of linear programming and complementarity as well as microeconomics.

Energy

Dr. Nesbitt's career spans the energy space worldwide. Dr. Nesbitt has worked on almost every major pipeline and LNG terminal that has entered since 1974; valued upstream, pipeline, and LNG assets in every area of North America and the world; valued Duke for the Duke-PanEnergy merger; developed power plants in a number of locations; calculated forward prices and price differentials in and between every North American electric generation region; supported landmark decisions such as the Bacton to Zeebrugge Interconnector, the TransMed pipeline(s), Maritimes and Northeast, and Ruby; helped win the famous Duke New Smyrna Beach case before the Florida PSC and the PGT Roll In case before the FERC; helped GRI administer its gas supply and utilization R&D portfolio; fully connected the prices of tradable emissions allowances with the price of energy; led the Calpine bankruptcy case in behalf of the Equity Committee and served as the economic witness in that contentious case; built the definitive world supply-transportation-refining-products-demand-pricing model of the oil business for one of the majors; and educated clients and managers in energy organizations worldwide. Dr. Nesbitt conceived the MarketPoint and ArrowHead energy modeling systems and built the economic centerpieces.

Other Industries

Outside energy, Dr. Nesbitt's consulting companies have done seminal work in high technology (product pricing, new product introduction), airlines (fleet allocation, scheduling, and seat pricing for United, Continental, SAS, British Air, KLM), rental car companies (Hertz), telecommunications (product line pricing for Pacific Bell, GTE, Bell Atlantic, NYNEX), transportation including trucks, ocean, and highway; infrastructure (various highway departments), resources (Anaconda-Arco, Potash Corporation of Saskatchewan), and environmental remediation. Dr. Nesbitt regularly briefs management and board level groups on energy and other commodity markets. Outside energy, Dr. Nesbitt cofounded and manages Reticle Inc. (developed and patented ultra-high surface area materials to purify water, plate metals, and make ultracapacitors). He cofounded Ferritech Inc. (patented biotechnology to manufacture ferric ions for use as an oxidant and dissolvent.)

History

Dr. Nesbitt began at the Xerox Palo Alto Research Center (PARC) as an analyst in the Management Systems Department (employee ID 70.) He moved to Stanford Research Institute (SRI) in 1974, where he worked on the seminal energy models the SRI-Gulf Model and the SRI World Energy Model and became Manager of the Decision Analysis--Energy. In 1977, Dr. Nesbitt cofounded the consulting company Decision Focus Incorporated (DFI), where he remained as a principal until 1995 when he founded Altos Management Partners to continue assisting energy, high technology, telecommunications, and other industries. At that time, he founded MarketPoint Inc. to develop world class forward market and decision support technology to serve industry and government. He sold Altos and MarketPoint in 2011 and initiated the development of ArrowHead, a deterministic and probabilistic economic modeling system and a series of models within it. Dr. Nesbitt is known in the energy industry for his market analysis, including the North American Regional Gas (NARG) model, the ArrowHead Global Gas Model (AGGM), the ArrowHead Global Oil Model (AGOM), the European Gas Model, the North American Regional Electricity Model, the North American Emissions Model, and the North American Coal Model. The market modeling methods developed by Dr. Nesbitt have been used for many of the North American and world energy companies in the oil, gas, electricity, coal, and emissions business.

Education

Dr. Nesbitt holds a B.S. degree in Engineering Science from the University of Nevada—Reno, an M.S. degree in Mechanical Engineering at Stanford University, and an M.S. and a Ph. D. degree in Engineering-Economic Systems (Ph. D. dissertation defended with honors) from Stanford University. Dr. Nesbitt has published in the economic, energy, and operations research literature. He was appointed at Stanford in Management Science and Engineering (MS&E) in 2013.

CONTENTS

LIST OF FIGURES

1 EXECUTIVE SUMMARY

Transportation and logistics are pivotally important to modeling world gas markets and world oil markets. The logistics and basis differentials between regions and the transportability between regions, within a single continent or across continents, are central to quantifying and understanding world gas and oil markets, prices, quantities, capacity additions, and assets in place; electricity prices and flows nodally and internodally around a transmission grid; short term storage such as electricity storage and its impact on price and quantity; longer-term storage such as gas, oil, and reservoir hydro storage and its impact on price and quantity; and a full range of logistically-oriented economic phenomena. Such quantification cannot possibly be accomplished without an efficient, detailed, general, accurate, temporal economic model of existing and prospective transportation and storage. And one cannot hope for an accurate model or answer unless there is a massively high degree of disaggregation of the transportation system. (Aggregate representations of transportation will not suffice.)

The objective of this report is to provide recommendations for the best design (including method and implementation) for meeting EIA's requirements for the logistics component of EIA's Global Hydrocarbon Supply Model (GHySMo). This report will focus on two ways to represent and solve transportation and logistics problems (and in reality spatial economic problems in general that transcend storage and logistics), and we are going to explicate both here in great detail. The first approach is monolithic global welfare maximization, expressed as a set of complementarity relationships. (Linear programming is a sub-problem of this approach.) In this report, we term this approach the "complementarity" approach, and we term its practitioners "complementarians." The second approach is network-oriented microeconomic equilibrium, derived from microeconomic textbooks that offer a pure agent-based focus. We term this approach "network microeconomics," and we term its practitioners "economists."

An important goal of this work is to demystify the so-called complementarity-linear programming-monolithic global welfare maximization approach, which in our view has been misunderstood at best and systematically obfuscated at worst. For years, linear programmers and complementarians have been pleading: "We are an economic solution too." They have been arguing that they are just as general as microeconomic network modeling and therefore should be deeded equal footing. The reader will see why such statements are simply not correct, both in the context of a spatial transportation grid and in fact in a full, spatially and technologically distributed economic system.

This analysis (which starts with complementarity because of its mathematical difficulty and frequent obfuscation) demonstrates the following:

1) Complementarity, as a method for solving spatial world or regional transportation-logistical economic problems (and in fact general network economic problems), is oversized, risky, complex, and redundant. EIA should avoid it for transportation and logistical modeling (and in fact for economic modeling altogether).

    a) Complementarity is intrinsically and inescapably derived from monolithic global welfare maximization using a Samuelsonian consumers-plus-producers-surplus approach.

Network microeconomic equilibrium rejects the notion of monolithic global welfare functions of any type from first principles, focusing instead on agents and their interactions.

b) Only if complementarity scrupulously and completely honors the Samuelson global welfare assumptions (including fully integrable demand functions, no welfare function components other than Samuelsonian consumers' and producers' surplus, and no constraints other than production functions and volumetric balances), does it represent a competitive spatial transportation market (competitive in the textbook economic sense).

c) The size of the complementarity representation is far too large and inefficient, spanning

   i) the "Cartesian product" of the number of market hubs times the number of transportation **plus** non-transportation activities in one's network for linear programming (in the static, single time point situation).

   ii) the "Cartesian product" of the number of hubs **plus** the number of links times the number of hubs **plus** the number of links for complementarity (in the static, single time point situation). Complementarity is far larger than linear programming.

   Both grow to colossally large numbers for real world transportation problems EIA requires, which must have thousands of links, hubs, and many forward time points spanning the world. Complementarity becomes horrendously, unworkably large. Both are massively larger than the dimensionality of network microeconomic equilibrium, which is the number of hubs in the model (no Cartesian product).

d) Once complementarity equations for a spatial competitive market have been written, those complementarity equations are always amenable and can and should be solved analytically, never numerically. **<u>The complementarity equations are fully amenable to analytic solution; there is absolutely no need for numerical solution</u>**. One never needs all of the GAMS-AMPL "send it to the solver" architectural complexity. One can solve the complementarity equations analytically. One need never solve complementarity equations numerically or deploy complementarity algorithms because the equations are amenable to analytic solution, after which the reduced system of analytic equations is solved much more simply and directly.

e) When one solves the complementarity equations analytically, one ends up with the microeconomic, agent-based, network equilibrium modeling equations, which are much smaller, more efficient, more direct, more reviewable, easier to solve, easier to organize, easier to visualize, more amenable to interpretation and communication, and immutably correct. The size of the network microeconomic problem is much smaller—only related to the number of hubs—dramatically smaller than the Cartesian product of hubs plus links times hubs plus links (which, for complementarity, is many times larger).

   <u>Simple substitution of production functions into the Kuhn-Tucker conditions eliminates altogether the need for a complementarity solution for transportation (and in fact the full model)</u>. Ponder what that means. In a transportation equilibrium problem, one does not need nor want complementarity. It is too big, too hard, and requires too much coding, all of which are unnecessary because the analytical solution solves and thereby eliminates the complementarity equations altogether. Network equilibrium gives the same answer (under the sufficiency conditions) using a much more simple, compact, and direct calculation.

2) EIA is far better off to directly deploy microeconomic, agent-based, network equilibrium modeling from the outset, knowing that it can be the ultimate result of analytical solution of complementarity equations. It avoids all the pitfalls and difficulties, and it contains no hint of a global welfare function.

   a) With network microeconomics, there is zero risk of "polluting" the solution by advertently (or inadvertently) departing from sufficiency conditions for monolithic global welfare maximization.

      i) Complementarity and linear programming, descended as they are from monolithic global welfare maximization, are subject to advertent or inadvertent abuse precisely because they are descended from monolithic global welfare maximization. That is not at all appealing for an organization such as EIA with broad public scrutiny and desiring accuracy and empirical validity.

      ii) If, using complementarity, one fools around with the welfare function or fools around with the constraints, one loses any and all connection from shadow price to price and any and all connection with market structure or operation. It is too easy to make errors, and it is too easy (and too tempting) to cheat. EIA (and other economic modelers) should avoid linear programming or complementarity to forestall such practice.

   b) The economic network that is the model is a very convenient way to organize and represent world or regional transportation structure. **<u>The picture of the model is the model</u>**. All EIA needs to do to build a transportation model is to build the picture of the model.

      i) Every economic node in the model, including but not limited to transportation, has a supply curve on its output.

      ii) Every economic node in the model, including but not limited to transportation, has a demand curve on its input.

      iii) Every hub in the model has a supply curve on every input link and therefore a supply stack comprising all the inputs to the hub.

      iv) Every hub in the model has a demand curve on every output link and therefore a demand stack comprising all the outputs from the hub.

      v) Therefore, every hub in the model has a simple supply-demand curve pair, easily solvable. Solving the simple, scalar supply-demand problem at every hub is well understood, straightforward, and reliable. Simple cobweb, fixed point, Newton's, or auction methods all work well. There is no need for colossally complicated “full rank” methods such as complementarity or linear programming.

c) When you “drag in a node and drop it into a network,” you are systematically inserting a demand curve on its output and a supply curve on its input. (Those curves are programmed into the node.) Dragging and dropping and crafting networks writes a full system of supply and demand equations without doing any programming. (The programming is already resident within the nodes.) All you have to do is insert node data to quickly build and modify a very large transportation model (or a small one).

d) With network microeconomics, there is no need for EIA to write computer code to build a model. With network microeconomics, all coding is pre-done in a high-level language, not “selfie” coding in Visual Basic (VB) or similar language. (EIA can and will write code if you wish to represent different production functions, different transportation loss functions, different supply curves, or different demand curves, but once the coding is associated with the nodes, it is finished. The nodes become “objects,” and the objects are super-reliable code/data bundles. (That is what “objects” are.) EIA will not need to code every time for every model.) EIA can therefore spend its time understanding markets, building structure, disaggregating transportation, assembling data, and solving problems, not coding. EIA is not a computer science R&D team; EIA is an economic modeling team.

e) With network microeconomics, EIA can easily represent transportation and economic phenomena that are otherwise very difficult to represent. For example, we show in this report how easy it is to represent issues such as take or pay contracts, hub-and-spoke versus point-to-point transportation (and in fact hybrids of both), market power using network equilibrium, and endogenous representation of storage. To do so with complementarity and Kuhn-Tucker conditions or linear programming would be markedly more difficult (perhaps impossible). Several sections of this report substantiate these issues mathematically.

f) EIA can represent storage as an endogenous player using the model summarized herein.[1] We have never seen that in any technique other than microeconomic network equilibrium. That is important in world and continental oil and gas and increasingly important in electricity, renewables, and the environment. An entire section is dedicated to the mathematics of storage.

g) The network microeconomic approach is algorithmically parallelizable, systematically parallelized rather than left to the serendipity of the compiler or the “solver” to hopefully schedule, advance, and thereby parallelize calculations. Network microeconomic equilibrium is uniquely amenable to systematic, non-serendipitous, scheduled, rational parallelization. The execution time is dramatically lower because of such parallelization. We show precisely why herein.

h) Microeconomic network equilibrium with a graphical interface is available to and runnable by everyone in a modeling organization, staff and contractors alike. There is no need to

[1] The storage model was validated by Dr. Jeffrey Williams, probably the world’s expert in commodity storage. Discussion and reference are included in the report.

concentrate the modeling within a few people who know where the "knobs" are and to rely on the ability and good will of those people to get the models built and run. Microeconomic transportation network models are imminently transparent to everyone, and everyone can contribute easily. The efficiency of modeling burgeons. Model structure, data, and architecture become "open architecture" to everyone. No staff member or group is on the critical path to getting EIA's mission accomplished to model transportation. The mathematics and simplicity of the approach laid out herein show why.

i) Subject matter experts (SMEs) are free to assemble network structure, gather data, review inputs and results, and suggest extensions and modifications. SMEs can work on subject matter. Modelers can easily implement that subject matter. Such division of labor is a boon to complex organizations such as EIA, who have to gather data and points of view and put them into models, and different people must do each. EIA cannot have algorithms, inflexibilities, rigidities, or ongoing equation writing or coding limiting your ability to build and operate detailed transportation and logistics models or interact with SMEs, which is what is likely to happen with linear programming and complementarity.

3) "Solvers" for complementarity (e.g., GAMS, AMPL) are concentrated, labor intensive, opaque (actually invisible), error prone to load, hugely oversized, likely size limited, "full rank", highly iterative, computer-architectural challenges of major proportion. EIA should avoid incorporating or using them.

a) Writing equations and "putting them into a solver" using FORTRAN-looking, line-oriented commands is 1970s-vintage computer science, extremely labor intensive and error prone. The GAMS code that appears in Gabriel et. al., presented as state of the art, may be acceptable for graduate students, but it is far from commercially or operationally acceptable to create or solve transportation networks in the fast paced, demanding EIA modeling environment. That type of code or coding is simply not acceptable in the modern world (even if it were needed, which it is not). EIA should be taken aback by that. This report shows how EIA can replace that with analytical solutions.

b) EIA cannot in any reasonable time add, delete, augment, amend, manage, run, interpret, etc. a complementarity-based model of world transportation. EIA has tried that with linear programming "solvers" for over two decades, and even that environment has proven cumbersome and opaque. EIA's cycle time has been longer than desirable. Microeconomic network modeling assuages that.

c) Solution methods for complementarity problems are iterative. Gabriel op. cit. asserts that they are multivariate Newton solvers, and the unreliability of multivariate Newton's methods is well known. They rely on iterative recalculation and inversion of a colossal-sized Jacobean matrix, whose size is equal to the "Cartesian product" of <u>the number of unknown prices plus the number of unknown quantities</u> times <u>the number of unknown prices plus the number of unknown quantities</u>. Newton algorithms have to repopulate, invert, and multiply a colossal but very sparse Jacobian matrix of this size on every iteration, and there is no guarantee that the inverse even exists of is calculable or that the iteration converges. Even though the matrix is sparse, the inverse is not, and the inversion of a matrix this size is susceptible to floating point roundoff/truncation error propagation. Multivariate Newton's methods (which complementarity relies on according to Gabriel op. cit.) are colossally large, highly local, prone to fall into "holes" that are not the answer (mentioned by Gabriel op. cit.), and extremely difficult to code and implement. Thankfully, analytical solutions obviate the need as the report shows mathematically.

d) Complementarity and linear programming require probably two orders of magnitude higher labor time than the graphical network microeconomic approach. The simplicity of the mathematics and the formulation, which we derive herein, show why.

e) The documentation for network microeconomic equilibrium lies in every advanced microeconomics text published. Such documentation is central to the economics field. All one has to do is reference the standard literature. That tends to be far better documentation than model manuals that attempt to establish a link from linear programming or complementarity to economics (or ignore the connection altogether) or who merely cite operations research methods.

4) Linear programming is not and cannot be a complete solution for transportation networks or economic networks. (We show precisely why mathematically—it is embedded within a larger, complete problem.) Linear programming is a partial solution, and only if production functions are Leontief (an undesirable assumption). The argument is subtle, and we deliver the math to prove it herein. (This argument is not opinion or idle conjecture; it is objectively mathematical as we show herein.)

   If complementarity is as good as it is advocated to be, it will and should replace linear programming completely in every application other than market economics. (It should not be used for market economics.)

   a) Linear programming is not needed if complementarity is as good as advertised.

   i) Linear programming is a limiting special case of complementarity.

   ii) If complementarity works, toss away linear programs and "go nonlinear" from the outset.

5) The transportation problem can be isolated and separated from the supply and demand problems. That means EIA can use one methodology (e.g., network microeconomic equilibrium) for the transportation model and another module or modules for the supply and demand models, a hybrid approach. There is no mandate to use one and only one technique spanning the entire system. That will be clear in the mathematics presented herein. The separation we prove here allows network microeconomic equilibrium for transportation and specialized supply and demand models.

6) We offer all the requisite mathematics and proofs to support the aforementioned findings. They are not idle conjecture; they are mathematically demonstrated. The reduction from a massive, complicated, labor-intensive, "feed-the-solver" complementarity or linear programming problem to a network microeconomic problem is well over a factor of four, in terms of equation

count, in the transportation models herein. The resulting solution algorithm is also far simpler, observable, less risky, and direct. The degree of simplification is even larger in fully temporal, dynamic problems. Two orders of magnitude reduction in cost and effort to EIA may well be a conservative estimate for formulating and solving network equilibrium problems focused wholly or in part on transportation and logistics.

7) We recommend EIA use network microeconomic equilibrium for its transportation modeling. It is direct, size-unlimited, local, orders of magnitude quicker, and free from conceptual and implementation errors that plague complementarity and linear programming. Equally importantly, it facilitates easy and immediate admission of multiple time points with inter-temporal profit maximization and endogenous capacity addition. The extension to multiple time points is simple and direct (in contrast with complementarity or linear programming.)

This report is intended to be **mathematically definitive and omit no steps**. The report is intended as the germ of a monograph on the subject of network microeconomic equilibrium versus complementarity (with linear programming as a special case of complementarity), for solving network-oriented economic equilibrium problems, which transportation problems always are. We are grateful for the sponsorship of EIA to initiate it, and we are looking to publish this as a monograph (as recommended by our reviewers). This document is intended to be definitive on the issue of complementarity versus network equilibrium. It shows mathematically and algorithmically the decisive size, efficiency, and operational advantages of microeconomic network modeling in the transportation and economic context. And when the sufficiency conditions hold, microeconomic techniques actually solve the complementarity problem. Microeconomics is a far simpler solution technique for the complementarity equations than complementary itself. While we have chosen a specific example or two to prove all the salient points, the discussions all generalize easily to the case of interconnected networks that contain more than transportation. (The results are not limited to any particular illustrative examples.) We intend this as a definitive document on economic modeling of transportation and logistics in a market setting.

For simplicity here, we focus on a strictly static formulation (implicitly serial static). To introduce the dynamic, multi-temporal dimension, the capital formation dimension, would complicate the notation. We leave that for a later day. We will state (without proof herein) that expansion to inherently dynamic representations with endogenous capacity expansion decisions even more strongly favors network microeconomic equilibrium. The compression in dimensionality is more dramatic when multiple time points enter the picture.

We have heard over the years several comments regarding the issues raised in this paper.

"I am just a model runner or manager. I don't really need to know the mathematics or the underlying concepts." Such assertions are preposterous; a Luddite point of view is not commensurate with managing or using a model. One does not need a Ph.D. in mathematical economics, but one needs a basic, fundamental understanding and internalization of the underlying methodology, economic science, and mathematics. Sure, one has to understand the real world, the industry, computer architecture, data gathering, data input, practicality, etc., but one absolutely cannot be an effective model builder, user, or manager without basic, intrinsic understanding.

"We have done just fine with linear programming. Why change?" Just because one has been running a particular modeling methodology for years does not sidestep the intrinsic shortcomings of that methodology. Assuredly EIA knows that linear programming is not up to the task of modeling world scale transportation or economic systems. If the method were up to the task, there would be no reason for this and similar papers. Has EIA really done all that well with linear programming? Is linear programming really suitable for the problems of the future (or the present)? Is linear programming, as certain vendors assert, truly a complete economic modeling approach? No; to ask the question is to answer it. Why would a different approach be contemplated if the old one were working? This paper is going to show exactly why linear programming is merely a sub-solution of part of the problem and is too big and ponderous to represent world scope transportation and process networks. When linear programming was chosen by EIA in the early 1990s, it may have been the only operational alternative available. Technology and methodology have moved beyond that in two and a half decades; network microeconomic equilibrium (and other methods) having since emerged that superseded it both methodologically and implementationally.

"If complementarity and network microeconomic give the same answer, hooray. We are vindicated in using monolithic global welfare maximization expressed as complementarity equations. Isn't that the way to go?" **Absolutely the opposite is true; complementarity is far and away the worst way to go by a long stretch, and this report will show why**. Complementarity is too big, indirect, unexplainable, prone to conceptual errors (welfare function, constraints), prone to programming errors ("delivering equations to a solver"), prone to data errors ("delivering data to a solver"), not amenable to decentralized algorithms, not amenable to user-friendly features, and intrinsically tied to 1970s-vintage computer science ("command lines to deliver information to solvers"). Complementarity jumps the gun to embrace numerical algorithms rather than first appealing to easily available and highly simplifying analytical solutions (solutions that are always available, as this report demonstrates). If one starts with complementarity, one can immediately substitute and solve the complementarity problem analytically so that complementarity itself is no longer needed.

Isn't microeconomic network equilibrium the same as GEMS or LEAP from the 1970s? Not in the slightest. Microeconomic equilibrium and its embodiments in MarketBuilder and ArrowHead and the methods discussed herein have no commonality with GEMS. The draft Leidos paper by Lauren K. Busch (see references) is incorrect regarding network equilibrium modeling and models in its discussion of MarketBuilder. Dale Nesbitt and Mark Nickeson wrote MarketBuilder/MarketPoint and later Alan Clark and Doug Kaweski contributed, all under Dale Nesbitt's direction. Dale Nesbitt conceptualized and caused MarketBuilder to be built to microeconomic textbook specifications as described herein. There is no methodological or algorithmic commonality between GEMS (which Dale Nesbitt and others built in the 1970s) and MarketBuilder, ArrowHead, or network microeconomic equilibrium. GEMS had economic flaws and was retired by the early 1990s. Nothing about GEMS extrapolates or applies to MarketBuilder, ArrowHead, or network microeconomic equilibrium; GEMS was an "evolutionary dead end." The significant point of departure started with Dale Nesbitt's 1984 **Operations Research** paper and the research that led to it. (A careful read of that paper shows that the method had already departed from GEMS/LEAP in favor of agent-based microeconomic equilibrium as early as 1984. It took more than another decade to get method and software completed.)

2 OUTLINE AND APPLICABILITY TO EIA REQUIREMENTS

This report is intended as a treatise on transportation and network modeling (and economic modeling in general), and it is also intended to be fully and completely responsive to the requirements that EIA established. We outline the report in the first subsection, and we summarize the EIA requirements in the second subsection and identify sections that contain the relevant discussion.

2.1 Outline and Organization of the Report

We begin in Section 3 by putting forth a simple yet representative model of a transportation grid, a transportation-oriented economic system.[2] Our logistical model interconnects four supply or export regions with four demand or import regions, embodying an interconnecting point-to-point type of transportation system. (We address other forms such as hub-and-spoke in a later section herein.) This simple example will serve as the basis for our mathematical and model developments. In order to be crystal clear how the complementarity method works, we pose a simple yet sophisticated example and formulate it using the complementarity approach. We formulate it thereafter using the network microeconomic approach. The fact that our demonstration is by example does not limit the generality of our mathematical proofs and developments. Our findings can be proven in the abstract as well, but abstraction is more difficult to follow. We want EIA to carefully understand the approaches, what flows from these approaches, what their limitations are, and why. Using the concrete example facilitates that goal.[3]

For the representative transportation example, we put forth in Section 4 a detailed explication of the monolithic global welfare maximization approach, from which the complementarity approach is directly derived.[4] The monolithic global welfare maximization approach encompasses both the complementarity approach and the linear programming approach (the latter shown to be a mere subset of the overall problem and not a complete solution in itself). We will demonstrate the extreme size and cumbersomeness of the complementarity approach, and we will show exactly why linear programming cannot by itself be a complete economic solution.[5] At best, linear programming can only be part of an economic solution. The purpose of this discussion is to show explicitly why complementarity and linear programming are so complex and oversized and work poorly for distributed transportation (and in fact economic network) problems. Part of the

[2] Everything developed in this paper pertains to networks containing transportation links and also every other process. Because the focus of this paper is transportation, we will emphasize that dimension. However, the arguments and mathematical developments generalize to networks comprised of process models including but not limited to transportation.

[3] One of the main oversights we perceive with complementarity is that there have not been concrete transportation or other problems that we have seen in the literature. Models have been abstract or conceptual. When one has a concrete, workable, palpable application, one can see exactly what the method is doing and more importantly why one always can and should use an analytical solution rather than a numerical solution in place of complementarity problems.

[4] We are not particularly enamored of monikers like MLP or any of the other acronyms for some specific kind of complementarity that is being invoked. We use the term complementarity very carefully to represent the expression of Kuhn-Tucker conditions for a fully nonlinear monolithic global welfare maximization problem. If all or part of it specializes to a linear or "mixed" complementarity relationship, that is no more than a special case. We are not interested in special cases; we are interested in generality.

[5] We thought that linear programming could be a complete solution in the days of the Brock-Nesbitt National Science Foundation report. We have since learned that it cannot.

discussion in Section 4 is to analyze one of the complementarity solution algorithms as a way to show concretely why the problem literally skyrockets in size and complexity.

We then in Section 5 put forth a detailed mathematical explication of the microeconomic network equilibrium approach, which **by construction** does not admit of any monolithic global welfare function or any system-wide maximization problem. There is no danger of misusing or misinterpreting a monolithic global welfare function because there isn't one. We shall see how powerful it is to ban by assumption any and all monolithic global welfare considerations. (One key benefit is that network microeconomics requires no arcane assumptions such as integrability of demand.) We develop the network microeconomic approach for the same transportation problem for which the complementarity approach was developed in Section 4 and show that the network microeconomics problem is less than ¼ of the size in the microeconomic equilibrium context, far more organized, far more graphical, far more conceptually elegant, and far easier to solve the selfsame equations that complementarity labors to solve. The simplicity of the microeconomic approach relative to the complex and cumbersome complementarity approach will be obvious. The beauty of microeconomics is that it is a **direct** approach. It rejects welfare functions and monolithic maximizations right from the outset and deals with agents, their incentives, and their actions directly. By rejecting problematic concepts, there is negligible risk of misuse or misinterpretation.

Dale Nesbitt and a colleague Dr. Horace "Woody" Brock wrote what we were told was a definitive methodological treatise in 1977. We were repeatedly asked to expand and produce it as a textbook on energy modeling. (It was used in a few university economics courses. We have provided it to EIA.) We never did so in spite of the fact that the report was considered definitive. This report is intended to be equally definitive. We demonstrate here concretely the excess size, complexity, and cumbersomeness of indirect solutions such as complementarity. We demonstrate that if one elects to solve transportation and other problems analytically, whether beginning with complementarity or not, there is no need for complementarity. There is available a far more simple, direct, and correct approach. While we have chosen specific examples to prove all the salient points, all those proofs generalize to the case of interconnected networks that contain more than transportation. The beauty of using the example to derive everything is that people can understand it; it is not abstract.

This analysis is long overdue in the transportation context (and outside the transportation context in the general economic modeling framework as well.) As far as we know, no one has put together the true and correct mathematical and architectural relationship in the context of a real transportation problem between monolithic global welfare maximization and agent-based equilibrium. Sure, there can be an equivalency between the two. **But who cares?** There is an equivalency between an abacus and a modern computer. But who wants the ponderousness, slowness, and labor intensity of the abacus? Microeconomics is a much simpler, more direct, smaller, more transparent, direct approach to the problem that does not require interpretation, theorems, sufficiency conditions, avoidance of welfare function issues, etc. It allows EIA to attain the highly disaggregated representation of transportation and logistics that it needs. We will show quite clearly that in cases in which there is an equivalency, there is a better, easier, and more efficient way to solve complementarity equations, and it isn't direct numerical complementarity algorithms. That is the hard way, not to mention an old-school, command line oriented, FORTRAN-looking, low computer science, low economic science way. The best way is to use analytical substitution to solve, and by so doing eliminate, the complementarity equations altogether and thereby end up with a dramatically simpler system. Who wouldn't want to deploy analytic solutions in place of numerical solutions? No one (save perhaps for vendors who want EIA to continue to pay to build and operate complementarity or linear programming models.) Analytics trumps numerics. This equivalency is articulated in Section 7. Think what this means. It means that you need not implement complementary equations at all to get the complementarity solution. You substitute and thereby solve the complementarity equations analytically and to get complementarity solutions.

Following the complementarity discussion and the network equilibrium discussion and their equivalency, we have included sections to address transportation and logistical issues of great importance to EIA, all of which are addressed using network microeconomic methods.

1. First, if we have a particular transportation-supply-demand model and we need to change it, how easy is it to effect such change using monolithic global welfare maximization versus network equilibrium? With complementarity, it is not easy. With network equilibrium, it is a few drags and drops and a few data entries. Section 6 elaborates.

2. Second, we deal with the issue of hub-and-spoke transportation versus point-to-point transportation. This distinction is of great importance for issues such as LNG tanker dispatch and exchange and many pipelines around the world. We show how to represent such transportation methods efficiently using network microeconomic methods. Section 8 elaborates.

3. Third, we deal with the issue of take or pay (i.e., mandatory take) contracts in the LNG and pipeline market. We show specifically how to represent any such contracts that may exist in a transportation network using network microeconomics and how to pancake those contracts through intermediate markets from a supplier to a customer class. Section 9 elaborates.

4. Fourth, we discuss the issue of market power and prospective monopoly/oligopoly or monopsony/oligopsony behavior in and around a transportation network. We show why the issue is nowhere as simple as certain analysts suggest because in general one must model rather than specify the net demand function facing agents who may have market power. Section 10 elaborates.

5. Fifth, we discuss the requirements of the storage module needed to properly and accurately represent commodity storage. We are aware of no other embedded, agent-based, endogenous models of storage, so this section will be elucidating. That discussion appears in Section 11.

6. Sixth, we outline how and why EIA needs to model the endogenous entry of new capacity. It is not sufficient to attempt to insert capacity exogenously or to estimate it using a satellite model. Is must be endogenized. Section 12 overviews.

7. Seventh, we discuss issues related to "historical validation" of a model. We conclude that the value is, in fact, rather low, but we make recommendations how to do it. Section 13 is relevant here.

8. Eighth, we overview the notion of vertical integration. Is it required or desirable to force collections of decision makers to act in a model as if they were a vertically integrated entity? We will show in Section 14 that, in fact, the answer is no and EIA is best off with a decentralized representation

9. Ninth, we recommend the embedded timing conventions for the EIA model in Section 15.

10. Tenth, we put forth a few recommendations related to data management. That discussion assumes that EIA deploys the network microeconomic equilibrium method. Other methods have more difficult data management challenges. See Section 16.

11. Eleventh, we put forth recommendations and a number of dos and don'ts related to implementation. That discussion appears in Section 17.

We considered whether to include a discussion of how to represent alternative tariffing/pricing structures for transportation and transmission—postage stamp rates, zonal rates, point-to-point rates, pancaked rates, etc. We elected not to, but the full range of transportation tariffing is available in network microeconomic equilibrium, including hybrids of those methods. Alternative rate structures can be extremely important in electricity and pipelines in which regulators often mandate postage stamp or zonal rates. The method has it covered, but we have not amplified herein.

We initiate this paper with monolithic global welfare maximization and its manifestation of Kuhn-Tucker conditions expressed as complementarity equations. Why start with complementarity (the non-recommended method)? Why so much focus on the complementarity approach, the approach that is not recommended? Why not start the paper with network microeconomic equilibrium, the recommended method? The reason is that we must develop a complete, comprehensive, cold, hard, unequivocal, unassailable mathematical and conceptual representation of the complementarity approach and its intrinsic underlying underpinnings. It is so arcane, complicated, and ill-understood that we have to articulate and demystify it. Only then can we see clearly the excessive size, complexity, and conceptual limitations of the formulation and the method. Only then will we see the dangers of monolithic global welfare maximization. Only then will we have a set of equations and concepts to compare and contrast with network microeconomic equilibrium. Only then can we enumerate and articulate each and every one of the equations and enumerate how they can be solved analytically and simplified, avoiding altogether the need for a complementarity algorithm. **We are not aware that the complementarity formulation of the transportation networks presented herein has ever been published or widely understood**. (We are fairly confident it has not been widely understood.) We must understand complementarity implicitly if we are to understand why the complementarity equations can and should be solved analytically, reducing the problem by so doing directly to the simpler and more direct network microeconomic equilibrium method. We must understand complementarity carefully if we are to understand why network microeconomic equilibrium is the best and most efficient way to solve complementarity itself. Network equilibrium de facto is a complementarity "solver," a new and effective way to achieve the solution.

2.2 EIA Requirements

This section reproduces the specific requirements established by EIA for this analysis. (The EIA requirements are reproduced verbatim in bold blue font.) We have endeavored herein to be responsive to each of the EIA requirements. Associated with each requirement, in the discussion in this section, we reference the section and findings in this report in which the specific EIA issue is addressed.

1. **Write a Component Design Report (CDR) for a model of natural gas and liquid fuels logistics (i.e. transportation and distribution) which can interact with models of production, processing, storage, and demand. The CDR will include (but not be limited to)**

   a. **General suggestions for the overall model design of GHySMo, including method as well as implementation.** This report concentrates in great detail on method and implementation, and the recommendations for method and implementation are strong and direct. We offer specific recommendations as well as a number of dos and don’ts in Section 17 related to implementation. Those dos and don’ts are reiterated throughout the report but are concentrated in Section 17 for easy access.
   b. **A template for logistical economic modeling driven by competing and complementary “supply chains,” which contain matrices of production, transportation, storage, and demand components.** The example around which the methodological discussions occur put forth in Section 3 addresses very carefully the supply chain notion of transportation. In fact the entire concept of a network is and was originally driven many years ago by the desire to have a systematic representation of every competing and complementary supply chain in the world (or in the subset of the world that is the focus of your model). There is really no equally effective way to represent supply chains rather than visual, graphical networks. The example put forth herein is quite an ideal illustration of how EIA should and will have to represent transportation and logistics. It shows where and how supply fits in, where and how transportation fits in, and where and how demand fits in. It shows how the entire supply chain of competing and complementary technologies fits in. We have included a specific discussion of storage in Section 11. Storage is difficult to model, and we know of no other endogenous storage methods. Section 11 indicates how storage would be inserted into a system such as that in Section 3.
   c. **Specific recommendations for how EIA should model transportation of natural gas and liquid fuels in GHySMo.** EIA should adopt the network microeconomic equilibrium modeling technique, whether by “build” or “buy,” to represent transportation and logistics. The bulk of this report shows why from a mathematical and user perspective. EIA should adopt a sophisticated graphical interface and a hierarchical regional representation. Those notions are addressed and supported throughout the report.
   d. **The interaction between storage and logistics, which is crucially important to modeling logistics; … an endogenous model of storage, which can pertain to gas, liquids, and electricity storage.** EIA is correct to assert that electricity storage cannot be ignored, and of course gas and oil storage have proven important in world and local

settings. We have a sophisticated, endogenous model of storage within a market setting, one that is fully nodal-spatial-locational. The ArrowHead storage model is located at every hub throughout the model, meaning it is highly spatial and represents market area storage, production area storage, intermediate location storage, and end-user storage as discrete entities. The gas storage model must buy commodity from the hub, inject it to storage, hold it, and sell it back to the hub, taking account of all losses and costs, continuously throughout the model forward horizon. Storage must occur nodally, i.e., locationally. Storage is inherently local, but its impacts cascade throughout the entire model. Section 11 outlines the storage module we recommend EIA build or buy.

e. **Economic methods of logistical modeling.** We have intentionally created in this report a textbook on complementarity, linear programming, and network microeconomic equilibrium. A textbook is exactly what is needed given the widespread confusion about what is really under the covers with complementarity, linear programming, and network microeconomic equilibrium. Some of the material in this report may have published before, but we believe much of it has not. We believe it is the seed for a textbook on economic modeling.

f. **Comparison to other methods of logistical modeling.** We give short shrift to non-economic models of logistics (e.g., input-output or spreadsheet models). They are ultimately arbitrary, and they are not suitable for an organization with the scope and responsibilities of EIA to understand and quantify worldwide markets and prices in an economic setting. There is no need to say more than that. There are only three economic approaches to transportation and logistics that we know of (complementarity, linear programming, and network microeconomic equilibrium), and we fully address all three and show why network microeconomic equilibrium dominates.

g. **Relationship to the economic literature on logistical and transportation matrix modeling.** Our storage model has been validated by Dr. Jeffrey Williams of the University of California at Davis for the California Energy Commission, perhaps the leading expert in the world at conceiving and modeling storage from an economic perspective.[6] The microeconomic and operations research references are comprehensive; we are aware of no better or more comprehensive references than Mas-Colell et. al. and Jehle and Reny et. al.

h. **Market power issues related to logistics – how to conceive, identify, and model them.** See Section 10 for a comprehensive discussion. Market power is an extremely difficult and subtle issue in a spatially distributed economic model, not nearly as simplified as the discussions in Gabriel op. cit. and other sources would have one believe. The reality is that one cannot deal with market power in a simplistic way outside the context of the entire world model. Market power in distributed, network-oriented system is a poorly solved or unsolved problem. EIA needs an extremely large competitive model related to the agent with alleged market power, and EIA needs to flex that model through a range of assumptions to discern implicit demand elasticity and competitive fringe supply elasticity facing that agent. Section 10 summarizes.

i. **Granularity (e.g., disaggregation versus aggregation of logistical elements) – how much does aggregation cost you in terms of accuracy?** It is not acceptable to aggregate transportation infrastructure. The level of aggregation implicit in the Leidos

[6] See Williams and Wright's definitive economic monograph on storage of commodities.

report attributed to EIA and to other models does not appear to us to be sufficient for North American gas models and certainly not for oil and product models. That level of aggregation extrapolated to the world is not going to be sufficient. Aggregation badly distorts price and flow projections. Samuelson's 1952 paper presages the difficulty of aggregation. Nesbitt and Scotcher's 2009 paper amplifies that. EIA needs as much disaggregation as you can possibly muster, literally thousands of transportation links spanning the world to represent pipelines, tankers, and transmission infrastructure for oil and gas and transmission infrastructure for electric power. The minute EIA aggregates transportation, you will have de facto aggregated both the supply region immediately upstream and the demand region immediately downstream from the aggregated transportation. By so doing, you have arbitrarily assumed that all that aggregate supply in the upstream region sees a single price and all that aggregate demand in the downstream region sees a single price. It is the aggregation of located prices that pollutes the predictability of the model. This paper addresses that. EIA must have a transportation methodology that allows a colossal degree of disaggregation, some of which will be point-to-point transportation and some of which will be hub-and-spoke transportation (also a potentially important issue). Some might be priced by a postage stamp method, and some might be priced point-to-point. The specific type of transportation structure is quite material to results. Section 8 lays out issues of point-to-point, hub-and-spoke, and aggregates of each.

j. **Mathematical ability/inability to link to other kinds of models (e.g., linear programming).** The report addresses this squarely. It shows mathematically in Section 4 that the transportation logistical economic submodel can be distinguished and separated from the supply and demand models. By so doing, EIA can implement its transportation-process network model using one methodology (and we show specifically why the methodology should not be linear programming or complementarity—oversized and unnecessary), and the demand and supply models can be special purpose supply and demand models that interface with the transportation and infrastructure submodel (which we show should be network microeconomic equilibrium.) This decomposition and melding of different models is discussed in detail in Section 4. This decomposition result is so strong that we have elected to discuss it in careful mathematical detail.

k. **Accommodation of different time frames (e.g., short-term run logistics where capacity entry is not an issue and longer-term run logistics where capacity entry is an issue).** We have incorporated a section that recommends temporality. We recommend that EIA implement a single modeling framework, a single methodology, that allows completely arbitrary forward time increments and completely arbitrary forward time horizon. An arbitrary number of forward time points with an arbitrary inter-time point spacing will allow EIA to use the same modeling system for all your models. All you will have to do is shorten the time points for short-term models and turn off the switch that adds capacity endogenously.

l. **Considerations of contracts (e.g., take or pay contracts) as they affect pipeline or LNG infrastructure. (Europe and Asia have take-or-pay contracts for some LNG and pipeline routes, and those contracts affect utilization of such transportation infrastructure.)** We have dedicated Section 9 to the issue of contracting. We show how network microeconomic equilibrium allows an elegant solution to the contracting problem. That solution impacts the "dispatch" of the pipeline and tanker system as dictated by contracts and how the rest of the system responds to that. The method marks those contracts to market to show how far out of the money they may or may not be to help SMEs and modelers to understand the potential viability and permanence of such contracts.

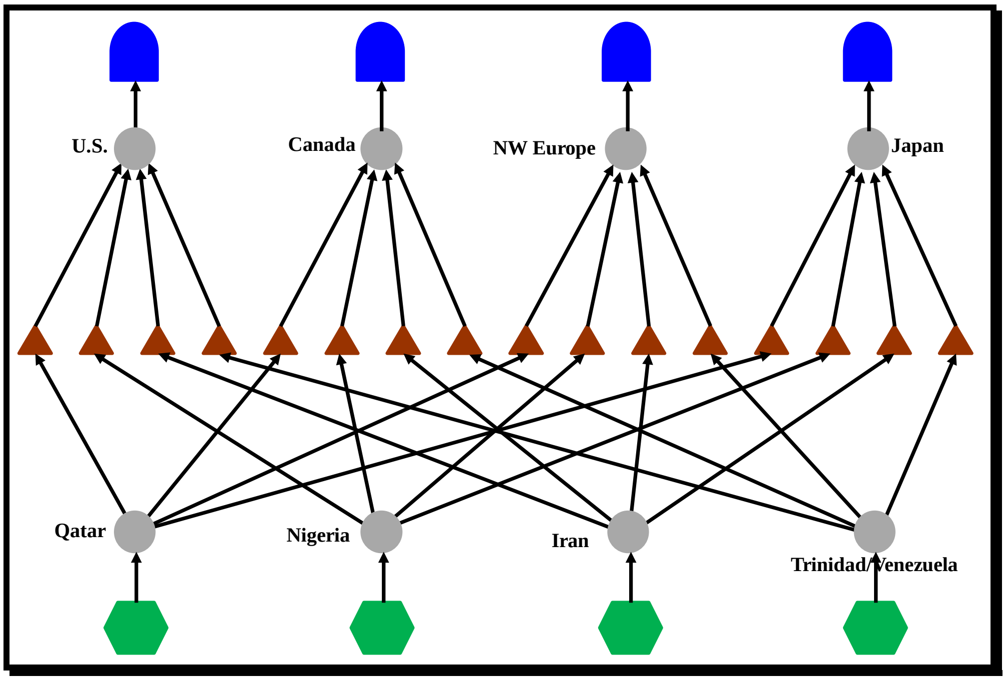
m. **Recommendations for computer architecture to implement the logistics model, including "do's" as well as "don'ts"** Section 17 discusses implementation issues in detail.
n. **Recommendations for data, data management, and knowledge management requirements for the logistical and storage sectors.** Section 16 discusses the data management and knowledge management issues.

A very large preponderance of this report is dedicated to the issue of methodology and implementation from a fundamental mathematical and economic point of view. That should be the primary criterion for EIA—method and efficiency of method. If you go wrong with method, it is impossible to recover. If you select complementarity or linear programming, you will go wrong with method.

3 A SIMPLE YET REPRESENTATIVE TRANSPORTATION/LOGISTICAL MODEL

For our purposes here, we are going to need a simple yet representative logistical transportation network for our analysis purposes. Consider the simple network in Figure 1, which is sufficiently small so as to be manageable yet sufficiently comprehensive to illustrate and embody most or all of the key concepts.[7]

Figure 1: A Sample Transportation Network



We articulate the dimensions of the simple transportation model in Figure 1 so that we can understand exactly how to expand the concepts here to the practical logistical modeling problem of:

- hundreds or thousands of new and existing gas pipeline links spanning a continent or continents plus a new and existing LNG tanker matrix spanning the world.
- hundreds or thousands of new and existing crude and product pipeline links spanning a continent or continents plus a new and existing crude, product, and VGO tanker matrix spanning the world.
- hundreds or thousands of new and existing electric transmission and distribution links spanning a region or a continent.

[7] This example is not intended to be comprehensive. In order to serve as a real world model for EIA, there will have to be dozens of supply regions and dozens of demand regions, all interconnected by a very large existing and prospective transportation system. This example is intended to illustrate the key points in a smaller context that can be easily extrapolated to the full world context.

In Section 4, we will implement the problem and solution as a monolithic global welfare maximization problem and write it as a complementarity problem in precisely the way Gabriel op. cit. did. Section 5 implements the same problem and solution as an agent-based general equilibrium model. The profound advantages of the latter will be immediately apparent.

Examining the diagram in Figure 1:

- There are four wellhead markets in the network—Qatar, Nigeria, Iran, and Trinidad/Venezuela. The wellhead markets are the circular hubs aligned horizontally at the second echelon from the bottom in Figure 1. Notice that local supply enters those markets, and those markets serve outbound competing transportation flows. (We use the notation of circular hubs for these wellhead markets. The circular hubs have one or more input flows and serve one or more output flows.) People have used the term "upstream" to characterize the supply side of the energy system. People think of the "upstream" as the green hexagonal supply nodes, and people think of the supply regions as "export" regions.

- Coming into each wellhead market hub at the second echelon from the bottom is the supply from each of the four supply regions at the first echelon at the bottom. Each regional supply is represented by a green primary resource supply hexagon (our convention, replaced by a green diamond in the new ArrowHead system), incoming from the bottom of the network. We see that the logistical equilibrium problem is not independent of the nature of supply in each region. To understand why, consider that if there is a paucity of supply in a given wellhead market, there will not be very much commodity to transport through any outbound transportation link; therefore those links will flow little or zero. The price in that region will be high, and exports from that region will not be competitive in downstream import regions. By contrast, if there is a surfeit of supply in a given wellhead market, there will be a large quantity of commodity to transport through the various outbound transportation links; therefore, those links will tend to flow at higher volumes. Prices will be low, and exports from that region will be highly competitive in downstream regions. Transportation flows are certainly not independent of upstream supplies.

- At the other end of the network (the top), there are four regional demand regions—United States, Canada, Northwest Europe, and Japan. They are represented by the circular hubs in the second echelon from the top in Figure 1. Notice that local, regional demand draws from those markets, and those markets serve inbound competing transportation flows. (We use the notation of circular hubs for these consumption markets. The circular hubs have one or more input flows and serve local markets.) The demand side of the system is often referred to as a "downstream" side.

- There are four demand curves, one for each of the demand regions, located in the tombstone nodes (our convention, replaced by blue squares in the new ArrowHead system) at the echelon at the top of the network. The demand curves, which represent local consumption in each of the regions, draw from the regional demand market hubs. In particular, each regional demand is represented by a blue demand tombstone (our convention), incoming from the regional markets. We shall see that the logistical equilibrium problem is not independent of the nature of demand in each region. To understand why, consider that if there is a paucity of demand in a given wellhead market, there will be little or no need or motivation to transport commodity through any inbound transportation link. The price will be low, and transporters will not profitably serve it. Thus, those links will flow little or zero. By contrast, if there is a large quantity of demand in a given regional market, there will be a large quantity of commodity "sucked in" through the various inbound transportation links and therefore those links will flow at higher volumes. Price will be high, and transporters will compete to serve it. Transportation flows are not independent of demand. People have used the term "downstream" to represent the demand side of the equation. To summarize, one of the functions of transportation is to fill these demand side hubs with commodity so that the local consumption can be serviced. People think of the demand regions or hubs as "import" regions.

- Once we have defined the supply regions or "export" regions and have defined the demand regions or "import" regions, the nature of the transportation logistical problem becomes apparent. In this example, we have four potential export regions and four potential import regions. That means there are potentially sixteen (four times four) point-to-point transportation links.[8] In real world modeling, it is not acceptable to consider only those transportation links that might be available and in place today. EIA's logistical model must consider emerging transportation links as well as existing transportation links. To be clear, in this example there are sixteen point-to-point transportation links, one originating from every supply region and terminating in every demand region. Collectively, these sixteen transportation links connect every supply region with every demand region. Competition among these transportation links determines inter-regional flows and basis differentials.

In summary, in this simple, representative example in Figure 1, there are:

- Four (4) demand nodes (four demand regions).
- Four (4) supply nodes (four supply regions).
- Sixteen (16) transportation nodes (all routes from all supply regions to all demand regions).

That means there are twenty-four (24) economic nodes in this model.

In addition, there are eight (8) hubs, four supply region hubs or supply region markets, and four demand region hubs or demand region markets.

The solution must balance inflows and outflows; the aggregate total of input volumes must be equal to the aggregate total of output volumes at each hub. Therefore, there are a total of 32 nodes in the model, 24 economic nodes and 8 market hubs.

Turning to the links, which is where the flows will actually be recorded, there are:

[8] We will discuss hub-and-spoke transportation configurations in a subsequent section. For now, we will focus on this point-to-point transportation model. There is no generality lost by this assumption.

- Four (4) links entering the demand nodes from the demand region hubs.
- Sixteen (16) links leaving the transportation nodes and entering the demand region hubs.
- Sixteen (16) links entering the transportation nodes after having left the regional wellhead hubs.
- Four (4) links leaving the supply nodes entering the regional wellhead hubs.

Thus there are a total of 40 links in the model, which means that the model in Figure 1 is comprised of 32 nodes (24 economic and 8 hub) and 40 links.

We can see how in real world transportation systems the number of transportation links compounds combinatorially. If one doubled the number of supply regions from 4 to 8 and doubled the number of demand regions from 4 to 8, one would see the number of transportation links increase from 16 to 64 in a point-to-point transportation model. We can see why real world transportation and logistics models escalate very rapidly in size and, in fact, become extremely large. They absolutely must be large as we have discussed with EIA because aggregation of a transportation matrix creates substantial problems with model predictive accuracy. Because of the sheer size explosion, EIA cannot afford difficult or ponderous formulations or formulations that restrict or hinder the number of transportation links that can realistically be implemented or calculated. That can be a very real limitation for techniques such as linear programming, nonlinear programming, or complementarity and a very real advantage for techniques like network microeconomic equilibrium.[9] EIA cannot live with methods that impose aggregation of the transportation and logistics segment.

We have also discussed and will reiterate here that aggregation of transportation links, transportation corridors, demand regions, or supply regions is a mortal difficulty for economic modeling of transportation and infrastructure. Aggregation obviates the ability to understand and predict markets, i.e., to understand prices or quantities throughout the world. Aggregation not only distorts the transportation matrix, but it also miscalculates supply and demand. Disaggregation is not a trivial issue. If the ponderousness of an algorithm or software solution is de facto forcing aggregation of an otherwise complex and detailed transportation matrix, then that algorithm and software system need to be eliminated from consideration. Any logistical model must of necessity be highly disaggregated if EIA is to have any hope of representing supply, production, wellhead price, transportation, end use price, or demand in a world or continental setting. We shall return to that theme repeatedly herein.[10] There appears a nice illustration and discussion of the perils of

[9] The ponderousness of linear programming, nonlinear programming, or complementarity is compounded exponentially when we consider multiple, interconnected time periods into the model. The temporal discussion can well be deferred until after we deal thoroughly with the spatial-transportation dimension.

[10] We understand that EIA's North American gas model may not have as disaggregated a representation of regional demand and the interregional transportation system EIA might like. The draft Leidos paper by Busch dated August 14, 2014, particularly Figure 1-1 contained therein, would seem to indicate that. Linear programming demands a substantially oversized and ponderous representation of highly spatial, highly disaggregated transportation systems and in our experience almost forces aggregation. We shall see why mathematically later in this report.

aggregation of supply regions, demand regions, and transportation infrastructure in Nesbitt and Scotcher.[11]

Aggregation is not a trivial point, but it cannot in practice be driven to the unrealistic extreme. One could not represent every last section of existing and prospective pipe in North America bigger than 2 inches in diameter. One ultimately is obliged to aggregate regions or groups of assets. The admonition here is that EIA needs a very high degree of disaggregation, the more the merrier, but ultimately one will have to stop disaggregating and work with small but reasonable aggregates. The severe admonition is that the clunkiness, ponderousness, or inefficiency of one's formulation or solution algorithm must never be what is forcing EIA to aggregate. Yet that is precisely what can happen with linear programming, nonlinear programming, and complementarity. This paper will show exactly why that is true from a mathematical perspective. It is not really a matter of opinion or judgment or opinion on the authors' part but rather mathematics and architecture.

[11] Nesbitt, Dale M. and Scotcher, Jill N., "Spatial Price and Quantity Relationships in World and Continental Commodity Markets," **The Energy Journal**, 2009 Special Issue.

4 BUILDING A LOGISTICAL MODEL USING COMPLEMENTARITY (WHICH IS DIRECTLY DERIVED FROM MONOLITHIC GLOBAL WELFARE MAXIMIZATION)

This section articulates in detail the approach of building a transportation/logistics model using a monolithic global welfare maximization or complementarity formulation.[12] The intent of this section is to articulate correctly and comprehensively **what one must assume in order to invoke a complementarity formulation**. What are the specific assumptions and interpretation one is required to make in order to obtain a set of complementarity equations to solve a network transportation problem? We are not aware of a preexisting full articulation such as this presented or referenced in the literature save for the famous Samuelson 1952 paper.[13] We have studied the cursory discussion of the complementarity approach to competitive markets in Gabriel op. cit., but the crucial details necessary to get analytically to the bottom of things are omitted.

Part of our intention is to put forth a mathematically precise articulation of the complementarity formulation of a transportation and logistics problem (and in fact by easy extension to an entire spatial economic problem.) **This has been long overdue**. Another part of our intention is to show precisely why linear programming is a special case and in fact a subset of the complementarity formulation. They are two cats from the same litter. People quip that complementarity is just "nonlinear linear programming!" That is not quite true, and we shall clarify the subtleties in this section. That is an important insight, and it presages why complementarity is more difficult and error prone to implement than linear programming by virtue of the sheer size increase and the need to deal with nonlinearity. As large and ponderous as the linear programming formulation gets, the complementarity approach is substantially larger and more difficult to solve because it embraces nonlinearity in places where linear programming gets by with linearity and because the size of the Jacobean matrix (which the complementarity solution algorithms need according to Gabriel op. cit.) is substantially larger than the size of the linear programming constraint matrix.

The section is intended to elucidate the economic assumptions one has to embrace, in the static temporal case, to appeal to the complementarity approach or to its close cousin linear and nonlinear programming. This section will emphasize the "brute force" nature of the approach, even for the simple example in Section 3. Extension to real world size transportation and logistical models with thousands of existing and prospective transportation activities (e.g., a world pipeline and tanker network) and with dozens of forward time points will be seen to be daunting and largely unworkable, particularly when compared to the network microeconomic equilibrium approach to the same problem articulated in Section 5.

---

[12] We consider only a static, single time point representation here. We defer the issue of market dynamics for a later time after dealing with the static, single time point representation.

[13] Samuelson, Paul A. (1952). "Spatial Price Equilibrium and Linear Programming," **The American Economic Review**, Vol. 42, No. 3 (Jun.), pp. 283-303. This paper is one of the best and most insightful papers in the annals of economics and an absolute must for economists in this field to fully internalize and understand. Our approach here is consistent with Samuelson's.

4.1 The "Complementarity" Approach (As Derived from Monolithic Global Welfare Maximization Kuhn-Tucker Conditions)

We begin in this section with the most detailed, most general, most comprehensive approach to complementarity for the example transportation problem in Section 3. Gabriel op. cit. have put together a nice monograph on complementarity and related topics. This section is consistent with their discussion, particularly the portions related to the formulation of the competitive economic transportation-conversion problem as a monolithic global welfare maximization problem.[14] Gabriel op. cit. is completely clear about that, and so are we. The complementarity formulation of a regional or world economic system revolves intrinsically and fundamentally around conceiving it as a monolithic global welfare maximization problem. It is significant to emphasize in these prefatory remarks that complementarity is an operations research approach, not an economic approach. Complementarity is not mentioned in any of the major graduate textbooks on microeconomics,[15][16][17] not even one single mention we could find. This cannot be because complementarity is a "newer" approach developed subsequent to those publications, a "high tech" approach such as an iPad that has only recently stormed onto the scene. Samuelson op. cit. was aware of complementarity in 1952 in his seminal paper on the topic. Indeed, complementarity and monolithic global welfare maximization have been around for many decades, all the way back to Marshall and his original conceptualization of consumers' surplus and Kuhn and Tucker, who proved the Kuhn-Tucker conditions. We reiterate that maximization of a monolithic global welfare function is not the technique of choice in microeconomics. Rather, microeconomics argues for individual, agent-based, behavioral models of producers and consumers developed and solved from first principles and interconnected by a network. We shall emphasize that point in the microeconomics section to follow (Section 5). We will take a step by step monolithic global welfare maximization approach to the transportation problem in Figure 1, which is quite general and representative of the types of transportation problems EIA will have to represent and solve, and it is easily expandable.

4.1.1 Step 1: Enumerate the Hubs (Alphabetic Designations)

To begin the complementarity formulation, we must designate and enumerate the hubs in the model. We designate them alphabetically from A through H as shown in Figure 2. We will need these hubs to enforce all the material balance points in the model. There is no escape from the need to letter or number or otherwise enumerate and point to the hubs in the model. As we shall see, the complementarity approach enumerates these hubs (linear programming doing so down the rows of a constraint matrix).

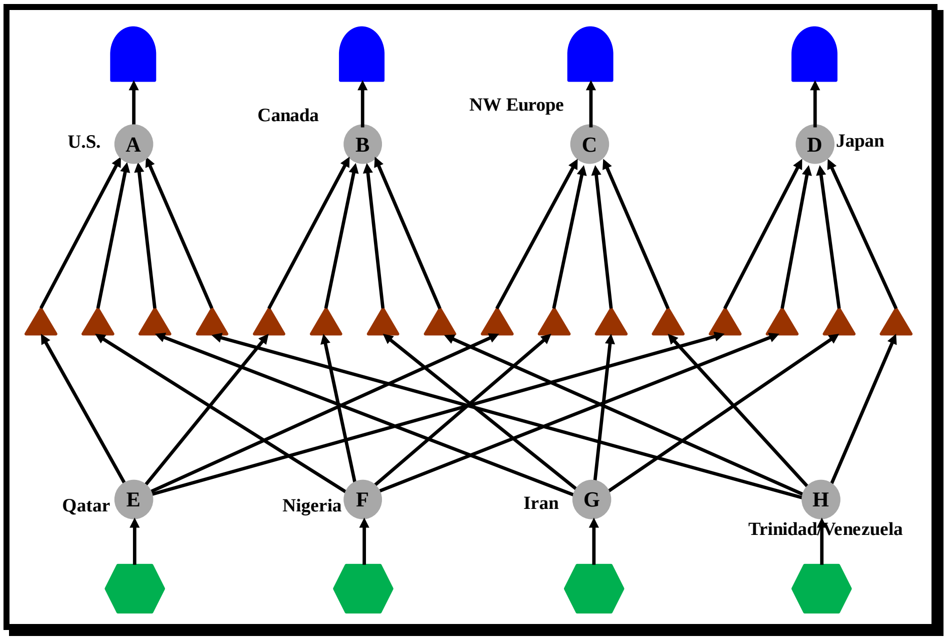
[14] Gabriel, Steven A., Conejo, Antonio J., Fuller, J. David, Hobbs, Benjamin F., and Ruiz, Carlos, **Complementarity Modeling in Energy Markets**, International Series in Operations Research and Management Science, Springer, 2013.

[15] Geoffrey A. Jehle and Philip J. Reny, **Advanced Microeconomic Theory**, (Third Edition), Prentice Hall, Apr 30, 2011.

[16] Andreu Mas-Colell, Michael D. Whinston and Jerry R. Green, **Microeconomic Theory**, Oxford University Press, Jun 15, 1995.

[17] Hal R. Varian, **Microeconomic Analysis**, Third Edition, Norton, Mar 17, 1992.

Figure 2: Network After Enumerating (Lettering) the Hubs



4.1.2 Step 2: Enumerate the Economic Nodes (Numeric Designations)

Similarly, we must designate and enumerate the economic activities in the model. We designate them numerically from 1 through 24. We will need these activities to enumerate cost and loss points in the model. See Figure 3. There is no escape from the need to enumerate the economic activities in the model. As we shall see, the complementarity approach enumerates these economic nodes. The linear programming approach does that along the "columns" of its constraint matrix. The Nobel laureate Koopmans and Hitchcock articulated this quite clearly for linear programming. We shall discuss the Koopmans-Hitchcock linear programming formulation later in this section.

4.1.3 Step 3: Enumerate and Index the Flows between Activities

Every link in the network represents a commodity flow. That is obvious, almost tautologically so, but it is nonetheless worthy of emphasis. It is convenient to index the quantities in terms of the origin point at the beginning of the link and the destination point at the end of the link where the arrowhead is. We have enumerated the indexed quantities in Figure 4. Notice that there is a single quantity on every link. After all, that is what the links signify—flow of a commodity from one node to another node. (One can anticipate that these quantities become time vectors when we generalize to the fully temporal case.)

Figure 3: Network After Enumerating (Numbering) the Economic Nodes

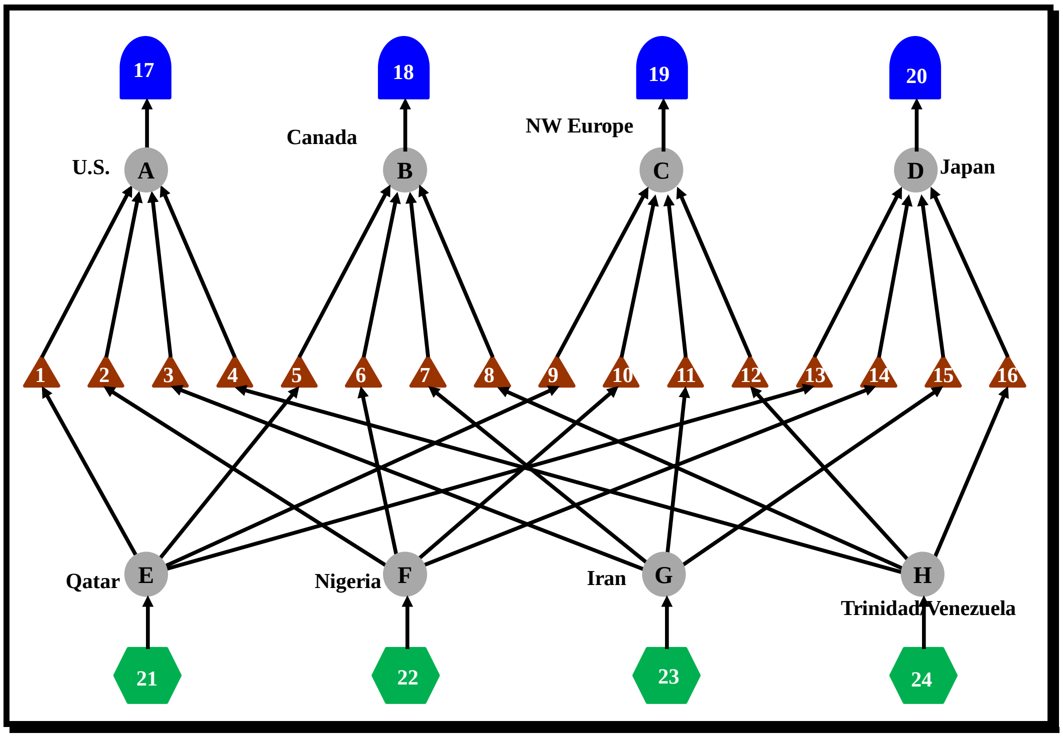
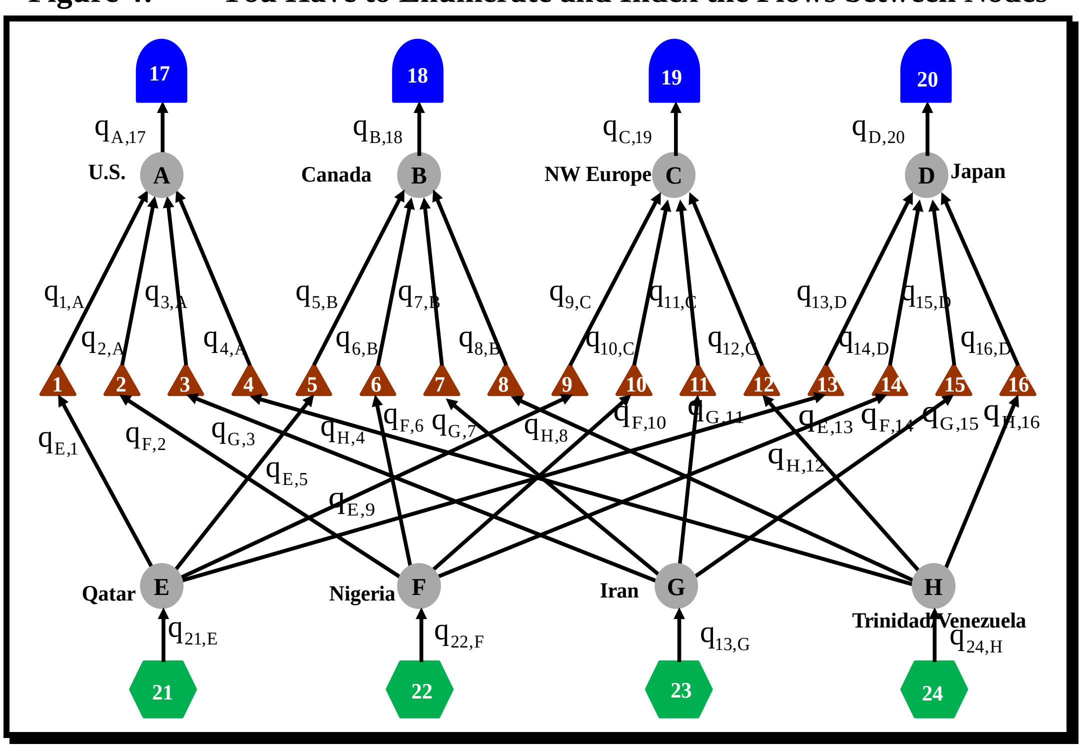


Figure 4: You Have to Enumerate and Index the Flows between Nodes



4.1.4 *Step 4: Specify Harberger-Marshallian-Samuelsonian Consumers' Surplus for Each of the Demand Nodes*

The first component of the monolithic global welfare function is the Harberger-Marshallian-Samuelsonian consumers' surplus. The Harberger-Marshallian-Samuelsonian consumers' surplus is the area under the demand curve (if such an area exists)[18] up to a given quantity of production. The areas under the four demand curves, assuming that they exist, in the transportation model in Figure 3 are the following:

$$CS_{17}\left(q_{A,17}\right) = \int_0^{q_{A,17}} p_{17}\left(x\right)dx$$

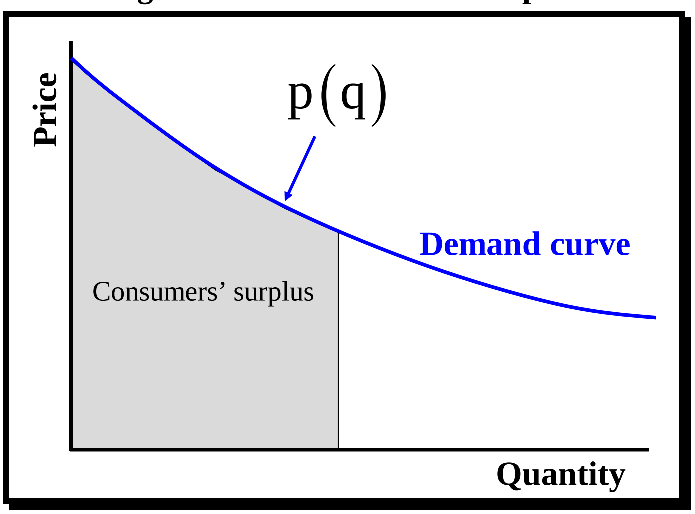
$$CS_{18}\left(q_{B,18}\right) = \int_0^{q_{B,18}} p_{18}\left(x\right)dx$$

$$CS_{19}\left(q_{C,19}\right) = \int_0^{q_{C,19}} p_{19}\left(x\right)dx$$

$$CS_{20}\left(q_{D,20}\right) = \int_0^{q_{D,20}} p_{20}\left(x\right)dx$$

These are the areas under the demand curves, one for every demand curve in the transportation model. Figure 5 illustrates these Marshallian consumers' surplus measures.

Figure 5: Consumers' Surplus



[18] Keep in mind, in real world models, demand curves are not usually integrable, and there is no requirement that they be integrable. Indeed, in practice, aggregate demand functions are rarely if ever integrable. There are more functional forms for aggregate demand curves that are not integrable than there are that are integrable. That means there is no area under the demand curve in practical models. However, it is instructive to proceed with the development assuming integrable demand curves.

Notice that the derivatives of these consumers' surplus functions are the indirect demand functions[19] themselves (assuming the demand functions are integrable), which can be proved by appealing to the fundamental theorem of the calculus.

$$\frac{\partial CS_{17}\left(q_{A,17}\right)}{\partial q_{A,17}} = p_{17}\left(q_{A,17}\right)$$
$$\frac{\partial CS_{18}\left(q_{B,18}\right)}{\partial q_{B,18}} = p_{18}\left(q_{B,18}\right)$$
$$\frac{\partial CS_{19}\left(q_{C,19}\right)}{\partial q_{C,19}} = p_{19}\left(q_{C,19}\right)$$
$$\frac{\partial CS_{20}\left(q_{D,20}\right)}{\partial q_{D,20}} = p_{20}\left(q_{D,20}\right)$$

Gabriel op. cit. mention the problems that arise from non-integrability of demand functions. We will not dwell on the integrability or lack of integrability of demand functions here in order to center our focus on logistics, but integrability of demand is a significant issue for monolithic global welfare maximization models. Developers of complementarity and monolithic global welfare measures generally ignore or assume away the issue, asserting that even though the solution method is developed based on strict assumptions of integrability, the solution can accommodate nonintegrable demand models. That is not technically true and extremely problematic; however, we are not going to pursue that avenue here. We continue by assuming integrability.

4.1.5 Step 5: Specify Total Cost for Each of the Supply Nodes

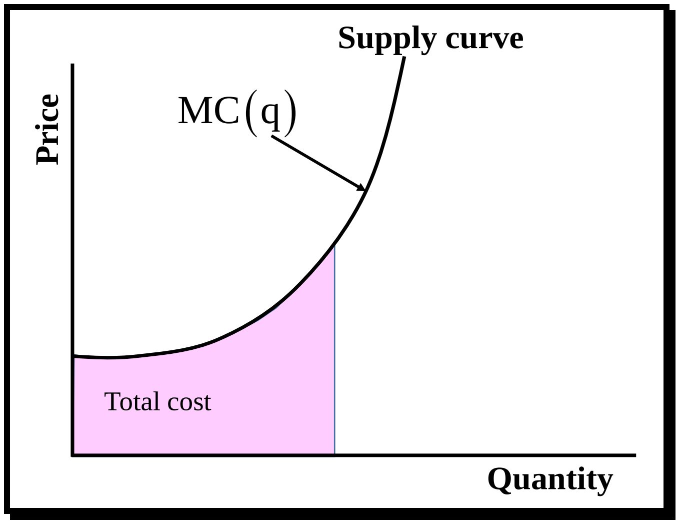
The total cost for each of the supply curves is the area under the supply curve out to the magnitude of production.

$$TC_{21}\left(q_{21,E}\right) = \int_0^{q_{21,E}} MC_{21}\left(x\right)dx$$
$$TC_{22}\left(q_{22,F}\right) = \int_0^{q_{22,F}} MC_{22}\left(x\right)dx$$
$$TC_{23}\left(q_{23,G}\right) = \int_0^{q_{23,G}} MC_{23}\left(x\right)dx$$
$$TC_{24}\left(q_{24,H}\right) = \int_0^{q_{24,H}} MC_{24}\left(x\right)dx$$

These are the total cost of supply areas as illustrated in Figure 6.

[19] A direct demand function expresses quantity consumed as a function of price. An indirect demand function, the inverse of a direct demand function, expresses price as a function of quantity consumed.

Figure 6: Total Cost of Supply



The derivative of the total cost curve is just the marginal cost curve itself, which we write for completeness as follows, which can be approved by appealing to the fundamental theorem of the calculus.

$$\frac{\partial TC_{21}\left(q_{E,21}\right)}{\partial q_{E,21}} = MC_{21}\left(q_{E,21}\right)$$
$$\frac{\partial TC_{22}\left(q_{F,22}\right)}{\partial q_{F,22}} = MC_{22}\left(q_{F,22}\right)$$
$$\frac{\partial TC_{23}\left(q_{G,23}\right)}{\partial q_{G,23}} = MC_{23}\left(q_{G,23}\right)$$
$$\frac{\partial TC_{24}\left(q_{H,24}\right)}{\partial q_{H,24}} = MC_{24}\left(q_{H,24}\right)$$

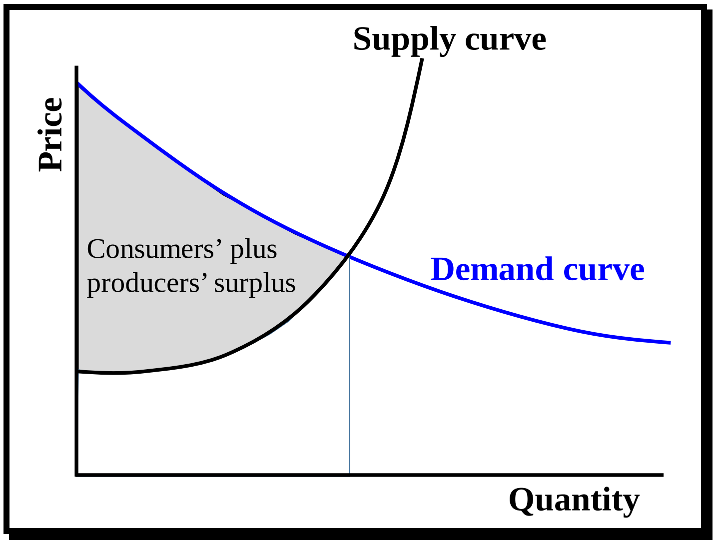
4.1.6 Step 6: Write Global Welfare as Consumers' Surplus minus Total Cost of Production minus Total Cost of Transportation

Using these concepts in the fashion that Samuelson op. cit. used them, we write the expression for monolithic global welfare using consumers' surplus minus transportation cost minus total cost of resource production, namely

$$\begin{aligned}
&CS_{17}\left(q_{A,17}\right)+CS_{18}\left(q_{B,18}\right)+CS_{19}\left(q_{C,19}\right)+CS_{20}\left(q_{D,20}\right)\\
&-VOC_1q_{1,A}-VOC_2q_{2,A}-VOC_3q_{3,A}-VOC_4q_{4,A}\\
&-VOC_5q_{5,B}-VOC_6q_{6,B}-VOC_7q_{7,B}-VOC_8q_{8,B}\\
&-VOC_9q_{9,C}-VOC_{10}q_{10,C}-VOC_{11}q_{11,C}-VOC_{12}q_{12,C}\\
&-VOC_{13}q_{13,D}-VOC_{14}q_{14,D}-VOC_{15}q_{15,D}-VOC_{16}q_{16,D}\\
&-TC_{21}\left(q_{21,E}\right)-TC_{22}\left(q_{22,F}\right)-TC_{23}\left(q_{23,G}\right)-TC_{24}\left(q_{24,H}\right)
\end{aligned}$$

This is akin to the notion of consumers' surplus plus producers' surplus in the economics literature. Consumers' plus producers' surplus is the area under the demand curve but above the supply curve, i.e., consumers' surplus minus total cost as in Figure 7. This area and its logical extension to include not only total upstream supply cost but also total transportation cost yields the monolithic global welfare function that is being maximized.

Figure 7: Consumers' Plus Producers' Surplus



In the complementarity formulations, there is no escape from this formulation—maximization of consumers' surplus less total cost. That concept lies at the fundamental heart of the complementarity approach. Gabriel op. cit. is crystal clear about that, as was Samuelson. We write the monolithic global welfare maximization problem in "Luenberger" sign convention format[20] to ensure that the signs of the Lagrange multipliers are scrupulously correct. (Luenberger adopts the convention that problems are expressed as minimum problems with less than or equal to constraints.)

$$
\begin{aligned}
\text{MIN}\quad & -CS_{17}\left(q_{A,17}\right)-CS_{18}\left(q_{B,18}\right)-CS_{19}\left(q_{C,19}\right)-CS_{20}\left(q_{D,20}\right)\\
& +VOC_1 q_{1,A}+VOC_2 q_{2,A}+VOC_3 q_{3,A}+VOC_4 q_{4,A}\\
& +VOC_5 q_{5,B}+VOC_6 q_{6,B}+VOC_7 q_{7,B}+VOC_8 q_{8,B}\\
& +VOC_9 q_{9,C}+VOC_{10} q_{10,C}+VOC_{11} q_{11,C}+VOC_{12} q_{12,C}\\
& +VOC_{13} q_{13,D}+VOC_{14} q_{14,D}+VOC_{15} q_{15,D}+VOC_{16} q_{16,D}\\
& +TC_{21}\left(q_{21,E}\right)+TC_{22}\left(q_{22,F}\right)+TC_{23}\left(q_{23,G}\right)+TC_{24}\left(q_{24,H}\right)
\end{aligned}
$$

[20] Luenberger, David G., **Linear and Nonlinear Programming**, Second Edition, Addison Wesley, May 1989. David G. Luenberger wrote a definitive text on linear and nonlinear optimization, and we use his sign conventions to ensure that the signs and directions of inequalities in our ultimate Kuhn-Tucker conditions are scrupulously correct.

4.1.7 Step 7: Apply the Losses for Each Transportation Process (Output-Input Function, i.e., Production Function)

We now turn to each of the sixteen transportation activities in the model. Each of the transportation links is characterized by a production function f(.), which can either be linear as in the Leontief case or nonlinear in the general case. Every production function for every transportation link is in principle different. These sixteen production function relationships (which are "output-input" or thermal efficiency relationships) are written:

$$
\begin{gathered}
q_{1,A} - f_1\left(q_{E,1}\right) = 0 \\
q_{2,A} - f_2\left(q_{F,2}\right) = 0 \\
q_{3,A} - f_3\left(q_{G,3}\right) = 0 \\
q_{4,A} - f_4\left(q_{H,4}\right) = 0 \\
q_{5,B} - f_5\left(q_{E,5}\right) = 0 \\
q_{6,B} - f_6\left(q_{F,6}\right) = 0 \\
q_{7,B} - f_7\left(q_{G,7}\right) = 0 \\
q_{8,B} - f_8\left(q_{H,8}\right) = 0 \\
q_{9,C} - f_9\left(q_{E,9}\right) = 0 \\
q_{10,C} - f_{10}\left(q_{F,10}\right) = 0 \\
q_{11,C} - f_{11}\left(q_{G,11}\right) = 0 \\
q_{12,C} - f_{12}\left(q_{H,12}\right) = 0 \\
q_{13,D} - f_{13}\left(q_{E,13}\right) = 0 \\
q_{14,D} - f_{14}\left(q_{F,14}\right) = 0 \\
q_{15,D} - f_{15}\left(q_{G,15}\right) = 0 \\
q_{16,D} - f_{16}\left(q_{H,16}\right) = 0
\end{gathered}
$$

We convert to an input-output function representation of transportation losses by simply inverting the production function or output-input curve. The input-output curve is an equivalent relationship to the output-input or thermal efficiency curve, merely its inverse (or reciprocal in the simplest cases).[21]

$$
\begin{gathered}
g_1\left(q_{1,A}\right) - q_{E,1} = 0 \\
g_2\left(q_{2,A}\right) - q_{F,2} = 0
\end{gathered}
$$

[21] We are not going to belabor inverses and inverse function theorems. For the types of production functions used here, they are invariably invertible. The entire development here works just as well with noninvertible production functions.

$$g_3\left(q_{3,A}\right)-q_{G,3}=0$$
$$g_4\left(q_{4,A}\right)-q_{H,4}=0$$
$$g_5\left(q_{5,B}\right)-q_{E,5}=0$$
$$g_6\left(q_{6,B}\right)-q_{F,6}=0$$
$$g_7\left(q_{7,B}\right)-q_{G,7}=0$$
$$g_8\left(q_{8,B}\right)-q_{H,8}=0$$
$$g_9\left(q_{9,C}\right)-q_{E,9}=0$$
$$g_{10}\left(q_{10,C}\right)-q_{F,10}=0$$
$$g_{11}\left(q_{11,C}\right)-q_{G,11}=0$$
$$g_{12}\left(q_{12,C}\right)-q_{H,12}=0$$
$$g_{13}\left(q_{13,D}\right)-q_{E,13}=0$$
$$g_{14}\left(q_{14,D}\right)-q_{F,14}=0$$
$$g_{15}\left(q_{15,D}\right)-q_{G,15}=0$$
$$g_{16}\left(q_{16,D}\right)-q_{H,16}=0$$

For the sample problem here, there are 16 input-output equations, one for each transportation node in the network. Thus, there will be 16 Lagrange multipliers on the sixteen input-output equation constraints.

4.1.8 Step 8: Balance Inputs and Outputs at Every Hub (Outputs minus Inputs Equals 0)

This set of equations will ensure that the inflow into each hub in the model is exactly equal to the output from each hub in the model.[22] Markets must "clear" in the sense that there is zero excess supply or zero excess demand at each of the market hubs.

$$q_{A,17}-q_{1,A}-q_{2,A}-q_{3,A}-q_{4,A}=0$$
$$q_{B,18}-q_{5,B}-q_{6,B}-q_{7,B}-q_{8,B}=0$$
$$q_{C,19}-q_{9,C}-q_{10,C}-q_{11,C}-q_{12,C}=0$$
$$q_{D,20}-q_{13,D}-q_{14,D}-q_{15,D}-q_{16,D}=0$$
$$q_{E,1}+q_{E,5}+q_{E,9}+q_{E,13}-q_{21,E}=0$$
$$q_{F,2}+q_{F,6}+q_{F,10}+q_{F,14}-q_{22,F}=0$$

[22] People have occasionally used ≥ 0 constraints to assert that inbound supply must be at least as large as outbound demand. Walrasian microeconomists reject that. We want the balance to be exact. We do not want to see free disposal going on at any node.

$$q_{G,3} + q_{G,7} + q_{G,11} + q_{G,15} - q_{23,G} = 0$$
$$q_{H,4} + q_{H,8} + q_{H,12} + q_{H,16} - q_{24,H} = 0$$

There are 8 balance equations at the hubs. That means there are 8 Lagrange multipliers at the hubs. Thus far we have 24 Lagrange multipliers, sixteen for the transportation input-output relationships and 8 for the hub quantity balances. These Lagrange multipliers on the quantity balances are going to be interpreted as prices. The Lagrange multipliers on those quantit balances is the sum total of all the Lagrange multipliers we are going to have for the problem except for the inequality constraints.

4.1.9 Step 9: Enumerate the Non-negativity Constraints Everywhere

Every quantity on every link must have a non-negativity constraint.

$$\begin{array}{llll} q_{1,A} \ge 0 & q_{2,A} \ge 0 & q_{3,A} \ge 0 & q_{4,A} \ge 0 \\ q_{5,B} \ge 0 & q_{6,B} \ge 0 & q_{7,B} \ge 0 & q_{8,B} \ge 0 \\ q_{9,C} \ge 0 & q_{10,C} \ge 0 & q_{11,C} \ge 0 & q_{12,C} \ge 0 \\ q_{13,D} \ge 0 & q_{14,D} \ge 0 & q_{15,D} \ge 0 & q_{16,D} \ge 0 \end{array}$$

$$\begin{array}{llll} q_{E,1} \ge 0 & q_{E,5} \ge 0 & q_{E,9} \ge 0 & q_{E,13} \ge 0 \\ q_{F,2} \ge 0 & q_{F,6} \ge 0 & q_{F,10} \ge 0 & q_{F,14} \ge 0 \\ q_{G,3} \ge 0 & q_{G,7} \ge 0 & q_{G,11} \ge 0 & q_{G,15} \ge 0 \\ q_{H,4} \ge 0 & q_{H,8} \ge 0 & q_{H,12} \ge 0 & q_{H,16} \ge 0 \end{array}$$

$$\begin{array}{llll} q_{A,17} \ge 0 & q_{B,18} \ge 0 & q_{C,10} \ge 0 & q_{D,20} \ge 0 \\ q_{21,E} \ge 0 & q_{22,F} \ge 0 & q_{23,G} \ge 0 & q_{24,H} \ge 0 \end{array}$$

4.1.10 Step 10: The Full Lagrangian

Using the objective function above, the input-output function constraints, and the hub balances, we can write the Lagrangian:

$$\begin{aligned} L = & -CS_{17}\left(q_{A,17}\right) - CS_{18}\left(q_{B,18}\right) - CS_{19}\left(q_{C,19}\right) - CS_{20}\left(q_{D,20}\right) \\ & + VOC_1 q_{1,A} + VOC_2 q_{2,A} + VOC_3 q_{3,A} + VOC_4 q_{4,A} + VOC_5 q_{5,B} + VOC_6 q_{6,B} + VOC_7 q_{7,B} + VOC_8 q_{8,B} \\ & + VOC_9 q_{9,C} + VOC_{10} q_{10,C} + VOC_{11} q_{11,C} + VOC_{12} q_{12,C} + VOC_{13} q_{13,D} + VOC_{14} q_{14,D} + VOC_{15} q_{15,D} + VOC_{16} q_{16,D} \\ & + TC_{21}\left(q_{21,E}\right) + TC_{22}\left(q_{22,F}\right) + TC_{23}\left(q_{23,G}\right) + TC_{24}\left(q_{24,H}\right) \\ & + \lambda_{A,17}\left[q_{A,17} - q_{1,A} - q_{2,A} - q_{3,A} - q_{4,A}\right] + \lambda_{B,18}\left[q_{B,18} - q_{5,B} - q_{6,B} - q_{7,B} - q_{8,B}\right] \\ & + \lambda_{C,19}\left[q_{C,19} - q_{9,C} - q_{10,C} - q_{11,C} - q_{12,C}\right] + \lambda_{D,20}\left[q_{D,20} - q_{13,D} - q_{14,D} - q_{15,D} - q_{16,D}\right] \end{aligned}$$

$$
\begin{aligned}
&+\nu_{E,1}\left[g_1\left(q_{1,A}\right)-q_{E,1}\right]+\nu_{F,2}\left[g_2\left(q_{2,A}\right)-q_{F,2}\right]+\nu_{G,3}\left[g_3\left(q_{3,A}\right)-q_{G,3}\right]+\nu_{H,4}\left[g_4\left(q_{4,A}\right)-q_{H,4}\right]\\
&+\nu_{E,5}\left[g_5\left(q_{5,B}\right)-q_{E,5}\right]+\nu_{F,6}\left[g_6\left(q_{6,B}\right)-q_{F,6}\right]+\nu_{G,7}\left[g_7\left(q_{7,B}\right)-q_{G,7}\right]+\nu_{H,8}\left[g_8\left(q_{8,B}\right)-q_{H,8}\right]\\
&+\nu_{E,9}\left[g_9\left(q_{9,C}\right)-q_{E,9}\right]+\nu_{F,10}\left[g_{10}\left(q_{10,C}\right)-q_{F,10}\right]+\nu_{G,11}\left[g_{11}\left(q_{11,C}\right)-q_{G,11}\right]+\nu_{H,12}\left[g_{12}\left(q_{12,C}\right)-q_{H,12}\right]\\
&+\nu_{E,13}\left[g_{13}\left(q_{13,D}\right)-q_{E,13}\right]+\nu_{F,14}\left[g_{14}\left(q_{14,D}\right)-q_{F,14}\right]+\nu_{G,15}\left[g_{15}\left(q_{15,D}\right)-q_{G,15}\right]+\nu_{H,16}\left[g_{16}\left(q_{16,D}\right)-q_{H,16}\right]\\
&+\lambda_{21,E}\left[q_{E,1}+q_{E,5}+q_{E,9}+q_{E,13}-q_{21,E}\right]+\lambda_{22,F}\left[q_{F,2}+q_{F,6}+q_{F,10}+q_{F,14}-q_{22,F}\right]\\
&+\lambda_{23,G}\left[q_{G,3}+q_{G,7}+q_{G,11}+q_{G,15}-q_{23,G}\right]+\lambda_{24,H}\left[q_{H,4}+q_{H,8}+q_{H,12}+q_{H,16}-q_{24,H}\right]\\
&-\mu_{1,A}q_{1,A}-\mu_{2,A}q_{2,A}-\mu_{3,A}q_{3,A}-\mu_{4,A}q_{4,A}-\mu_{5,B}q_{5,B}-\mu_{6,B}q_{6,B}-\mu_{7,B}q_{7,B}-\mu_{8,B}q_{8,B}\\
&-\mu_{9,C}q_{9,C}-\mu_{10,C}q_{10,C}-\mu_{11,C}q_{11,C}-\mu_{12,C}q_{12,C}-\mu_{13,D}q_{13,D}-\mu_{14,D}q_{14,D}-\mu_{15,D}q_{15,D}-\mu_{16,D}q_{16,D}\\
&-\mu_{E,1}q_{E,1}-\mu_{E,5}q_{E,5}-\mu_{E,9}q_{E,9}-\mu_{E,13}q_{E,13}-\mu_{F,2}q_{F,2}-\mu_{F,6}q_{F,6}-\mu_{F,10}q_{F,10}-\mu_{F,14}q_{F,14}\\
&-\mu_{G,3}q_{G,3}-\mu_{G,7}q_{G,7}-\mu_{G,11}q_{G,11}-\mu_{G,15}q_{G,15}-\mu_{H,4}q_{H,4}-\mu_{H,8}q_{H,8}-\mu_{H,12}q_{H,12}-\mu_{H,16}q_{H,16}\\
&-\mu_{A,17}q_{A,17}-\mu_{B,18}q_{B,18}-\mu_{C,19}q_{C,19}-\mu_{D,20}q_{D,20}-\mu_{21,E}q_{21,E}-\mu_{22,F}q_{22,F}-\mu_{23,G}q_{23,G}-\mu_{24,H}q_{24,H}
\end{aligned}
$$

We reiterate there are 40 flowing quantities (the links in the network). There are 8 Lagrange multipliers on the balance constraints. There are 16 Lagrange multipliers on the input-output equations. That gives a total of 64 unknowns exclusive of the Lagrange multipliers on the inequality constraints.

4.1.11 Step 11: Compute the Kuhn-Tucker Conditions for This Monolithic World Welfare Maximization

This section enumerates the Kuhn-Tucker conditions.

4.1.11.1 Derivatives of Demand Equations (Four)

Here are the four derivatives of the consumers' surpluses:

$$
\begin{aligned}
\frac{\partial L}{\partial q_{A,17}} &= -p_{17}\left(q_{A,17}\right)+\lambda_{A,17}-\mu_{A,17}=0\\
\frac{\partial L}{\partial q_{B,18}} &= -p_{18}\left(q_{B,18}\right)+\lambda_{B,18}-\mu_{B,18}=0\\
\frac{\partial L}{\partial q_{C,19}} &= -p_{19}\left(q_{C,19}\right)+\lambda_{C,19}-\mu_{C,19}=0\\
\frac{\partial L}{\partial q_{D,20}} &= -p_{20}\left(q_{D,20}\right)+\lambda_{D,20}-\mu_{D,20}=0
\end{aligned}
$$

So far, there are four equations.

4.1.11.2 Derivatives of Supply Equations (Four)

Following are the four derivatives of the supply equations:

$$\frac{\partial L}{\partial q_{21,E}} = MC_{21}\left(q_{21,E}\right) - \lambda_{21,E} - \mu_{21,E} = 0$$
$$\frac{\partial L}{\partial q_{22,F}} = MC_{22}\left(q_{22,F}\right) - \lambda_{22,F} - \mu_{22,F} = 0$$
$$\frac{\partial L}{\partial q_{23,G}} = MC_{23}\left(q_{23,G}\right) - \lambda_{23,G} - \mu_{23,G} = 0$$
$$\frac{\partial L}{\partial q_{24,H}} = MC_{24}\left(q_{24,H}\right) - \lambda_{24,H} - \mu_{24,H} = 0$$

So far, there are eight equations.

4.1.11.3 Derivatives of the Outputs of the Transportation Activities (Sixteen)

There are sixteen derivatives of outputs from transportation activities:

$$\frac{\partial L}{\partial q_{1,A}} = VOC_1 + \nu_{E,1} g_1{}'\left(q_{1,A}\right) - \lambda_{A,17} - \mu_{1,A} = 0$$
$$\frac{\partial L}{\partial q_{2,A}} = VOC_2 + \nu_{F,2} g_2{}'\left(q_{2,A}\right) - \lambda_{A,17} - \mu_{2,A} = 0$$
$$\frac{\partial L}{\partial q_{3,A}} = VOC_3 + \nu_{G,3} g_3{}'\left(q_{3,A}\right) - \lambda_{A,17} - \mu_{3,A} = 0$$
$$\frac{\partial L}{\partial q_{4,A}} = VOC_4 + \nu_{H,4} g_4{}'\left(q_{4,A}\right) - \lambda_{A,17} - \mu_{4,A} = 0$$
$$\frac{\partial L}{\partial q_{5,B}} = VOC_5 + \nu_{E,5} g_5{}'\left(q_{5,B}\right) - \lambda_{B,18} - \mu_{5,B} = 0$$
$$\frac{\partial L}{\partial q_{6,B}} = VOC_6 + \nu_{F,6} g_6{}'\left(q_{6,B}\right) - \lambda_{B,18} - \mu_{6,B} = 0$$
$$\frac{\partial L}{\partial q_{7,B}} = VOC_7 + \nu_{G,7} g_7{}'\left(q_{7,B}\right) - \lambda_{B,18} - \mu_{7,B} = 0$$
$$\frac{\partial L}{\partial q_{8,B}} = VOC_8 + \nu_{H,8} g_8{}'\left(q_{8,B}\right) - \lambda_{B,18} - \mu_{8,B} = 0$$
$$\frac{\partial L}{\partial q_{9,C}} = VOC_9 + \nu_{E,9} g_9{}'\left(q_{9,C}\right) - \lambda_{C,19} - \mu_{9,C} = 0$$
$$\frac{\partial L}{\partial q_{10,C}} = VOC_{10} + \nu_{F,10} g_{10}{}'\left(q_{10,C}\right) - \lambda_{C,19} - \mu_{10,C} = 0$$
$$\frac{\partial L}{\partial q_{11,C}} = VOC_{11} + \nu_{G,11} g_{11}{}'\left(q_{11,C}\right) - \lambda_{C,19} - \mu_{11,C} = 0$$

$$\frac{\partial L}{\partial q_{12,C}} = VOC_{12} + \nu_{H,12} g_{12}'(q_{12,C}) - \lambda_{C,19} - \mu_{12,C} = 0$$
$$\frac{\partial L}{\partial q_{13,D}} = VOC_{13} + \nu_{E,13} g_{13}'(q_{13,D}) - \lambda_{D,20} - \mu_{13,D} = 0$$
$$\frac{\partial L}{\partial q_{14,D}} = VOC_{14} + \nu_{F,14} g_{14}'(q_{14,D}) - \lambda_{D,20} - \mu_{14,D} = 0$$
$$\frac{\partial L}{\partial q_{15,D}} = VOC_{15} + \nu_{G,15} g_{15}'(q_{15,D}) - \lambda_{D,20} - \mu_{15,D} = 0$$
$$\frac{\partial L}{\partial q_{16,D}} = VOC_{16} + \nu_{H,16} g_{16}'(q_{16,D}) - \lambda_{D,20} - \mu_{16,D} = 0$$

Sixteen equations are added, bringing the total to twenty-four equations thus far.

4.1.11.4 Derivatives of the Inputs to the Transportation Activities (Sixteen)

There are sixteen derivatives of inputs to transportation activities:

$$\frac{\partial L}{\partial q_{E,1}} = -\nu_{E,1} + \lambda_{21,E} - \mu_{E,1} = 0$$
$$\frac{\partial L}{\partial q_{F,2}} = -\nu_{F,2} + \lambda_{22,F} - \mu_{F,2} = 0$$
$$\frac{\partial L}{\partial q_{G,3}} = -\nu_{G,3} + \lambda_{23,G} - \mu_{G,3} = 0$$
$$\frac{\partial L}{\partial q_{H,4}} = -\nu_{H,4} + \lambda_{24,H} - \mu_{H,4} = 0$$
$$\frac{\partial L}{\partial q_{E,5}} = -\nu_{E,5} + \lambda_{21,E} - \mu_{E,5} = 0$$
$$\frac{\partial L}{\partial q_{F,6}} = -\nu_{F,6} + \lambda_{22,F} - \mu_{F,6} = 0$$
$$\frac{\partial L}{\partial q_{G,7}} = -\nu_{G,7} + \lambda_{23,G} - \mu_{G,7} = 0$$
$$\frac{\partial L}{\partial q_{H,8}} = -\nu_{H,8} + \lambda_{24,H} - \mu_{H,8} = 0$$
$$\frac{\partial L}{\partial q_{E,9}} = -\nu_{E,9} + \lambda_{21,E} - \mu_{E,9} = 0$$
$$\frac{\partial L}{\partial q_{F,10}} = -\nu_{F,10} + \lambda_{22,F} - \mu_{F,10} = 0$$
$$\frac{\partial L}{\partial q_{G,11}} = -\nu_{G,11} + \lambda_{23,G} - \mu_{G,11} = 0$$
$$\frac{\partial L}{\partial q_{H,12}} = -\nu_{H,12} + \lambda_{24,H} - \mu_{H,12} = 0$$
$$\frac{\partial L}{\partial q_{E,13}} = -\nu_{E,13} + \lambda_{21,E} - \mu_{E,13} = 0$$

$$\frac{\partial L}{\partial q_{F,14}} = -\nu_{F,14} + \lambda_{22,F} - \mu_{F,14} = 0$$
$$\frac{\partial L}{\partial q_{G,15}} = -\nu_{G,15} + \lambda_{23,G} - \mu_{G,15} = 0$$
$$\frac{\partial L}{\partial q_{H,16}} = -\nu_{H,16} + \lambda_{24,H} - \mu_{H,16} = 0$$

Sixteen more equations are added, bringing the total thus far to forty equations.

4.1.11.5 Derivatives of the Lagrangian With Respect to the Lagrange Multipliers λ (Eight)

The balance equations are the derivative of the Lagrangian with respect to the Lagrange multipliers λ:

$$\frac{\partial L}{\partial \lambda_{A,17}} = q_{A,17} - q_{1,A} - q_{2,A} - q_{3,A} - q_{4,A} = 0$$
$$\frac{\partial L}{\partial \lambda_{B,18}} = q_{B,18} - q_{5,B} - q_{6,B} - q_{7,B} - q_{8,B} = 0$$
$$\frac{\partial L}{\partial \lambda_{C,19}} = q_{C,19} - q_{9,C} - q_{10,C} - q_{11,C} - q_{12,C} = 0$$
$$\frac{\partial L}{\partial \lambda_{D,20}} = q_{D,20} - q_{13,D} - q_{14,D} - q_{15,D} - q_{16,D} = 0$$
$$\frac{\partial L}{\partial \lambda_{21,E}} = q_{E,1} + q_{E,5} + q_{E,9} + q_{E,13} - q_{21,E} = 0$$
$$\frac{\partial L}{\partial \lambda_{21,E}} = q_{E,1} + q_{E,5} + q_{E,9} + q_{E,13} - q_{21,E} = 0$$
$$\frac{\partial L}{\partial \lambda_{22,F}} = q_{F,2} + q_{F,6} + q_{F,10} + q_{F,14} - q_{22,F} = 0$$
$$\frac{\partial L}{\partial \lambda_{23,G}} = q_{G,3} + q_{G,7} + q_{G,11} + q_{G,15} - q_{23,G} = 0$$
$$\frac{\partial L}{\partial \lambda_{24,H}} = q_{H,4} + q_{H,8} + q_{H,12} + q_{H,16} - q_{24,H} = 0$$

This adds eight equations. So far, there are forty-eight equations.

4.1.11.6 Derivatives of the Lagrangian with Respect to the Lagrange Multipliers ν (Sixteen)

The input-output equations are the derivative of the Lagrangian with respect to the Lagrange multipliers ν:

$$\frac{\partial L}{\partial \nu_{E,1}} = g_1\left(q_{1,A}\right) - q_{E,1} = 0$$

$$\frac{\partial L}{\partial \nu_{F,2}} = g_2\left(q_{2,A}\right) - q_{F,2} = 0$$

$$\frac{\partial L}{\partial \nu_{G,3}} = g_3\left(q_{3,A}\right) - q_{G,3} = 0$$

$$\frac{\partial L}{\partial \nu_{H,4}} = g_4\left(q_{4,A}\right) - q_{H,4} = 0$$

$$\frac{\partial L}{\partial \nu_{E,5}} = g_5\left(q_{5,B}\right) - q_{E,5} = 0$$

$$\frac{\partial L}{\partial \nu_{F,6}} = g_6\left(q_{6,B}\right) - q_{F,6} = 0$$

$$\frac{\partial L}{\partial \nu_{G,7}} = g_7\left(q_{7,B}\right) - q_{G,7} = 0$$

$$\frac{\partial L}{\partial \nu_{H,8}} = g_8\left(q_{8,B}\right) - q_{H,8} = 0$$

$$\frac{\partial L}{\partial \nu_{E,9}} = g_9\left(q_{9,C}\right) - q_{E,9} = 0$$

$$\frac{\partial L}{\partial \nu_{F,10}} = g_{10}\left(q_{10,C}\right) - q_{F,10} = 0$$

$$\frac{\partial L}{\partial \nu_{G,11}} = g_{11}\left(q_{11,C}\right) - q_{G,11} = 0$$

$$\frac{\partial L}{\partial \nu_{H,12}} = g_{12}\left(q_{12,C}\right) - q_{H,12} = 0$$

$$\frac{\partial L}{\partial \nu_{E,13}} = g_{13}\left(q_{13,D}\right) - q_{E,13} = 0$$

$$\frac{\partial L}{\partial \nu_{F,14}} = g_{14}\left(q_{14,D}\right) - q_{F,14} = 0$$

$$\frac{\partial L}{\partial \nu_{G,15}} = g_{15}\left(q_{15,D}\right) - q_{G,15} = 0$$

$$\frac{\partial L}{\partial \nu_{H,16}} = g_{16}\left(q_{16,D}\right) - q_{H,16} = 0$$

After adding sixteen equations, the total so far is sixty-four equations.

4.1.11.7 Kuhn-Tucker Conditions for Inequality Constraints

$$\mu_{1,A}q_{1,A}=0 \quad \mu_{1,A}\geq 0 \quad q_{1,A}\geq 0$$
$$\mu_{2,A}q_{2,A}=0 \quad \mu_{2,A}\geq 0 \quad q_{2,A}\geq 0$$
$$\mu_{3,A}q_{3,A}=0 \quad \mu_{3,A}\geq 0 \quad q_{3,A}\geq 0$$
$$\mu_{4,A}q_{4,A}=0 \quad \mu_{4,A}\geq 0 \quad q_{4,A}\geq 0$$
$$\mu_{5,B}q_{5,B}=0 \quad \mu_{5,B}\geq 0 \quad q_{5,B}\geq 0$$
$$\mu_{6,B}q_{6,B}=0 \quad \mu_{6,B}\geq 0 \quad q_{6,B}\geq 0$$
$$\mu_{7,B}q_{7,B}=0 \quad \mu_{7,B}\geq 0 \quad q_{7,B}\geq 0$$
$$\mu_{8,B}q_{8,B}=0 \quad \mu_{8,B}\geq 0 \quad q_{8,B}\geq 0$$
$$\mu_{9,C}q_{9,C}=0 \quad \mu_{9,C}\geq 0 \quad q_{9,C}\geq 0$$
$$\mu_{10,C}q_{10,C}=0 \quad \mu_{10,C}\geq 0 \quad q_{10,C}\geq 0$$
$$\mu_{11,C}q_{11,C}=0 \quad \mu_{11,C}\geq 0 \quad q_{11,C}\geq 0$$
$$\mu_{12,C}q_{12,C}=0 \quad \mu_{12,C}\geq 0 \quad q_{12,C}\geq 0$$
$$\mu_{13,D}q_{13,D}=0 \quad \mu_{13,D}\geq 0 \quad q_{13,D}\geq 0$$
$$\mu_{14,D}q_{14,D}=0 \quad \mu_{14,D}\geq 0 \quad q_{14,D}\geq 0$$
$$\mu_{15,D}q_{15,D}=0 \quad \mu_{15,D}\geq 0 \quad q_{15,D}\geq 0$$
$$\mu_{16,D}q_{16,D}=0 \quad \mu_{16,D}\geq 0 \quad q_{16,D}\geq 0$$
$$\mu_{E,1}q_{E,1}=0 \quad \mu_{E,1}\geq 0 \quad q_{E,1}\geq 0$$
$$\mu_{E,5}q_{E,5}=0 \quad \mu_{E,5}\geq 0 \quad q_{E,5}\geq 0$$
$$\mu_{E,9}q_{E,9}=0 \quad \mu_{E,9}\geq 0 \quad q_{E,9}\geq 0$$
$$\mu_{E,13}q_{E,13}=0 \quad \mu_{E,13}\geq 0 \quad q_{E,13}\geq 0$$
$$\mu_{F,2}q_{F,2}=0 \quad \mu_{F,2}\geq 0 \quad q_{F,2}\geq 0$$
$$\mu_{F,6}q_{F,6}=0 \quad \mu_{F,6}\geq 0 \quad q_{F,6}\geq 0$$
$$\mu_{F,10}q_{F,10}=0 \quad \mu_{F,10}\geq 0 \quad q_{F,10}\geq 0$$
$$\mu_{F,14}q_{F,14}=0 \quad \mu_{F,14}\geq 0 \quad q_{F,14}\geq 0$$
$$\mu_{G,3}q_{G,3}=0 \quad \mu_{G,3}\geq 0 \quad q_{G,3}\geq 0$$
$$\mu_{G,7}q_{G,7}=0 \quad \mu_{G,7}\geq 0 \quad q_{G,7}\geq 0$$
$$\mu_{G,11}q_{G,11}=0 \quad \mu_{G,11}\geq 0 \quad q_{G,11}\geq 0$$
$$\mu_{G,15}q_{G,15}=0 \quad \mu_{G,15}\geq 0 \quad q_{G,15}\geq 0$$
$$\mu_{H,4}q_{H,4}=0 \quad \mu_{H,4}\geq 0 \quad q_{H,4}\geq 0$$
$$\mu_{H,8}q_{H,8}=0 \quad \mu_{H,8}\geq 0 \quad q_{H,8}\geq 0$$
$$\mu_{H,12}q_{H,12}=0 \quad \mu_{H,12}\geq 0 \quad q_{H,12}\geq 0$$
$$\mu_{H,16}q_{H,16}=0 \quad \mu_{H,16}\geq 0 \quad q_{H,16}\geq 0$$
$$\mu_{A,17}q_{A,17}=0 \quad \mu_{A,17}\geq 0 \quad q_{A,17}\geq 0$$
$$\mu_{B,18}q_{B,18}=0 \quad \mu_{B,18}\geq 0 \quad q_{B,18}\geq 0$$
$$\mu_{C,19}q_{C,19}=0 \quad \mu_{C,19}\geq 0 \quad q_{C,19}\geq 0$$
$$\mu_{D,20}q_{D,20}=0 \quad \mu_{D,20}\geq 0 \quad q_{D,20}\geq 0$$
$$\mu_{21,E}q_{21,E}=0 \quad \mu_{21,E}\geq 0 \quad q_{21,E}\geq 0$$
$$\mu_{22,F}q_{22,F}=0 \quad \mu_{22,F}\geq 0 \quad q_{22,F}\geq 0$$
$$\mu_{23,G}q_{23,G}=0 \quad \mu_{23,G}\geq 0 \quad q_{23,G}\geq 0$$

$$\mu_{24,H} q_{24,H} = 0 \quad \mu_{24,H} \geq 0 \quad q_{24,H} \geq 0$$

4.1.12 *Step 12: Incorporate the Forty Derivatives into the Forty Inequality Constraints to Write the Forty Complementarity Relationships*

This section enumerates the Kuhn-Tucker conditions in complementarity form. They are created by substituting the foregoing derivatives into the inequality equations. In particularly, we isolate the Lagrange multipliers μ and then substitute them into the Kuhn-Tucker conditions to create the complementarity equations. This is the standard procedure, and it eliminates the Lagrange multipliers μ.

4.1.12.1 Four Demand Equations (4 Total)

$$\left[\lambda_{A,17} - p_{17}\left(q_{A,17}\right)\right] q_{A,17} = 0 \qquad \left[\lambda_{A,17} - p_{17}\left(q_{A,17}\right)\right] \geq 0 \qquad q_{A,17} \geq 0$$
$$\left[\lambda_{B,18} - p_{18}\left(q_{B,18}\right)\right] q_{B,18} = 0 \qquad \left[\lambda_{B,18} - p_{18}\left(q_{B,18}\right)\right] \geq 0 \qquad q_{B,18} \geq 0$$
$$\left[\lambda_{C,19} - p_{19}\left(q_{C,19}\right)\right] q_{C,19} = 0 \qquad \left[\lambda_{C,19} - p_{19}\left(q_{C,19}\right)\right] \geq 0 \qquad q_{C,19} \geq 0$$
$$\left[\lambda_{D,20} - p_{20}\left(q_{D,20}\right)\right] q_{D,20} = 0 \qquad \left[\lambda_{D,20} - p_{20}\left(q_{D,20}\right)\right] \geq 0 \qquad q_{D,20} \geq 0$$

4.1.12.2 Four Supply Equations (4 Total)

$$\left[MC_{21}\left(q_{21,E}\right) - \lambda_{21,E}\right] q_{21,E} = 0 \qquad \left[MC_{21}\left(q_{21,E}\right) - \lambda_{21,E}\right] \geq 0 \qquad q_{21,E} \geq 0$$
$$\left[MC_{22}\left(q_{22,F}\right) - \lambda_{22,F}\right] q_{22,F} = 0 \qquad \left[MC_{22}\left(q_{22,F}\right) - \lambda_{22,F}\right] \geq 0 \qquad q_{22,F} \geq 0$$
$$\left[MC_{23}\left(q_{23,G}\right) - \lambda_{23,G}\right] q_{23,G} = 0 \qquad \left[MC_{23}\left(q_{23,G}\right) - \lambda_{23,G}\right] \geq 0 \qquad q_{23,G} \geq 0$$
$$\left[MC_{24}\left(q_{24,H}\right) - \lambda_{24,H}\right] q_{24,H} = 0 \qquad \left[MC_{24}\left(q_{24,H}\right) - \lambda_{24,H}\right] \geq 0 \qquad q_{24,H} \geq 0$$

4.1.12.3 Sixteen Pairs of Transportation Relationships (32 Total)

$$\left[VOC_1 + \nu_{E,1} g_1'\left(q_{1,A}\right) - \lambda_{A,17}\right] q_{1,A} = 0 \qquad \left[VOC_1 + \nu_{E,1} g_1'\left(q_{1,A}\right) - \lambda_{A,17}\right] \geq 0 \qquad q_{1,A} \geq 0$$
$$\left(-\nu_{E,1} + \lambda_{21,E}\right) q_{E,1} = 0 \qquad \left(-\nu_{E,1} + \lambda_{21,E}\right) \geq 0 \qquad q_{E,1} \geq 0$$

$$\left[VOC_2 + \nu_{F,2} g_2'\left(q_{2,A}\right) - \lambda_{A,17}\right] q_{2,A} = 0 \qquad \left[VOC_2 + \nu_{F,2} g_2'\left(q_{2,A}\right) - \lambda_{A,17}\right] \geq 0 \qquad q_{2,A} \geq 0$$
$$\left(-\nu_{F,2} + \lambda_{22,F}\right) q_{F,2} = 0 \qquad \left(-\nu_{F,2} + \lambda_{22,F}\right) \geq 0 \qquad q_{F,2} \geq 0$$

$$\left[VOC_3 + \nu_{G,3} g_3'\left(q_{3,A}\right) - \lambda_{A,17}\right] q_{3,A} = 0 \qquad \left[VOC_3 + \nu_{G,3} g_3'\left(q_{3,A}\right) - \lambda_{A,17}\right] \geq 0 \qquad q_{3,A} \geq 0$$
$$\left(-\nu_{G,3} + \lambda_{23,G}\right) q_{G,3} = 0 \qquad \left(-\nu_{G,3} + \lambda_{23,G}\right) \geq 0 \qquad q_{G,3} \geq 0$$

$$\left[VOC_4 + \nu_{H,4} g_4'\left(q_{4,A}\right) - \lambda_{A,17}\right] q_{4,A} = 0 \qquad \left[VOC_4 + \nu_{H,4} g_4'\left(q_{4,A}\right) - \lambda_{A,17}\right] \geq 0 \qquad q_{4,A} \geq 0$$

$$\left(-\nu_{H,4}+\lambda_{24,H}\right)q_{H,4}=0 \qquad \left(-\nu_{H,4}+\lambda_{24,H}\right)\geq 0 \qquad q_{H,4}\geq 0$$

$$\left[VOC_5+\nu_{E,5}g_5'\left(q_{5,B}\right)-\lambda_{B,18}\right]q_{5,B}=0 \qquad \left[VOC_5+\nu_{E,5}g_5'\left(q_{5,B}\right)-\lambda_{B,18}\right]\geq 0 \qquad q_{5,B}\geq 0$$

$$\left(-\nu_{E,5}+\lambda_{21,E}\right)q_{E,5}=0 \qquad \left(-\nu_{E,5}+\lambda_{21,E}\right)\geq 0 \qquad q_{E,5}\geq 0$$

$$\left[VOC_6+\nu_{F,6}g_6'\left(q_{6,B}\right)-\lambda_{B,18}\right]q_{6,B}=0 \qquad \left[VOC_6+\nu_{F,6}g_6'\left(q_{6,B}\right)-\lambda_{B,18}\right]\geq 0 \qquad q_{6,B}\geq 0$$

$$\left(-\nu_{F,6}+\lambda_{22,F}\right)q_{F,6}=0 \qquad \left(-\nu_{F,6}+\lambda_{22,F}\right)\geq 0 \qquad q_{F,6}\geq 0$$

$$\left[VOC_7+\nu_{G,7}g_7'\left(q_{7,B}\right)-\lambda_{B,18}\right]q_{7,B}=0 \qquad \left[VOC_7+\nu_{G,7}g_7'\left(q_{7,B}\right)-\lambda_{B,18}\right]\geq 0 \qquad q_{7,B}\geq 0$$

$$\left(-\nu_{G,7}+\lambda_{23,G}\right)q_{G,7}=0 \qquad \left(-\nu_{G,7}+\lambda_{23,G}\right)\geq 0 \qquad q_{G,7}\geq 0$$

$$\left[VOC_8+\nu_{H,8}g_8'\left(q_{8,B}\right)-\lambda_{B,18}\right]q_{8,B}=0 \qquad \left[VOC_8+\nu_{H,8}g_8'\left(q_{8,B}\right)-\lambda_{B,18}\right]\geq 0 \qquad q_{8,B}\geq 0$$

$$\left(-\nu_{H,8}+\lambda_{24,H}\right)q_{H,8}=0 \qquad \left(-\nu_{H,8}+\lambda_{24,H}\right)\geq 0 \qquad q_{H,8}\geq 0$$

$$\left[VOC_9+\nu_{E,9}g_9'\left(q_{9,C}\right)-\lambda_{C,19}\right]q_{9,C}=0 \qquad \left[VOC_9+\nu_{E,9}g_9'\left(q_{9,C}\right)-\lambda_{C,19}\right]\geq 0 \qquad q_{9,C}\geq 0$$

$$\left(-\nu_{E,9}+\lambda_{21,E}\right)q_{E,9}=0 \qquad \left(-\nu_{E,9}+\lambda_{21,E}\right)\geq 0 \qquad q_{E,9}\geq 0$$

$$\left[VOC_{10}+\nu_{F,10}g_{10}'\left(q_{10,C}\right)-\lambda_{C,19}\right]q_{10,C}=0 \qquad \left[VOC_{10}+\nu_{F,10}g_{10}'\left(q_{10,C}\right)-\lambda_{C,19}\right]\geq 0 \qquad q_{10,C}\geq 0$$

$$\left(-\nu_{F,10}+\lambda_{22,F}\right)q_{F,10}=0 \qquad \left(-\nu_{F,10}+\lambda_{22,F}\right)\geq 0 \qquad q_{F,10}\geq 0$$

$$\left[VOC_{11}+\nu_{G,11}g_{11}'\left(q_{11,C}\right)-\lambda_{C,19}\right]q_{11,C}=0 \qquad \left[VOC_{11}+\nu_{G,11}g_{11}'\left(q_{11,C}\right)-\lambda_{C,19}\right]\geq 0 \qquad q_{11,C}\geq 0$$

$$\left(-\nu_{G,11}+\lambda_{23,G}\right)q_{G,11}=0 \qquad \left(-\nu_{G,11}+\lambda_{23,G}\right)\geq 0 \qquad q_{G,11}\geq 0$$

$$\left[VOC_{12}+\nu_{H,12}g_{12}'\left(q_{12,C}\right)-\lambda_{C,19}\right]q_{12,C}=0 \qquad \left[VOC_{12}+\nu_{H,12}g_{12}'\left(q_{12,C}\right)-\lambda_{C,19}\right]\geq 0 \qquad q_{12,C}\geq 0$$

$$\left(-\nu_{H,12}+\lambda_{24,H}\right)q_{H,12}=0 \qquad \left(-\nu_{H,12}+\lambda_{24,H}\right)\geq 0 \qquad q_{H,12}\geq 0$$

$$\left[VOC_{13}+\nu_{E,13}g_{13}'\left(q_{13,D}\right)-\lambda_{D,20}\right]q_{13,D}=0 \qquad \left[VOC_{13}+\nu_{E,13}g_{13}'\left(q_{13,D}\right)-\lambda_{D,20}\right]\geq 0 \qquad q_{13,D}\geq 0$$

$$\left(-\nu_{E,13}+\lambda_{21,E}\right)q_{E,13}=0 \qquad \left(-\nu_{E,13}+\lambda_{21,E}\right)\geq 0 \qquad q_{E,13}\geq 0$$

$$\left[VOC_{14}+\nu_{F,14}g_{14}'\left(q_{14,D}\right)-\lambda_{D,20}\right]q_{14,D}=0 \qquad \left[VOC_{14}+\nu_{F,14}g_{14}'\left(q_{14,D}\right)-\lambda_{D,20}\right]\geq 0 \qquad q_{14,D}\geq 0$$

$$\left(-\nu_{F,14}+\lambda_{22,F}\right)q_{F,14}=0 \qquad \left(-\nu_{F,14}+\lambda_{22,F}\right)\geq 0 \qquad q_{F,14}\geq 0$$

$$\left[VOC_{15}+\nu_{G,15}g_{15}'\left(q_{15,D}\right)-\lambda_{D,20}\right]q_{15,D}=0 \qquad \left[VOC_{15}+\nu_{G,15}g_{15}'\left(q_{15,D}\right)-\lambda_{D,20}\right]\geq 0 \qquad q_{15,D}\geq 0$$

$$\left(-\nu_{G,15}+\lambda_{23,G}\right)q_{G,15}=0 \qquad \left(-\nu_{G,15}+\lambda_{23,G}\right)\geq 0 \qquad q_{G,15}\geq 0$$

$$\left[VOC_{16}+\nu_{H,16}g_{16}'\left(q_{16,D}\right)-\lambda_{D,20}\right]q_{16,D}=0 \qquad \left[VOC_{16}+\nu_{H,16}g_{16}'\left(q_{16,D}\right)-\lambda_{D,20}\right]\geq 0 \qquad q_{16,D}\geq 0$$
$$\left(-\nu_{H,16}+\lambda_{24,H}\right)q_{H,16}=0 \qquad \left(-\nu_{H,16}+\lambda_{24,H}\right)\geq 0 \qquad q_{H,16}\geq 0$$

4.1.12.4 Sixteen Input-output Equations

The input-output equations are the derivatives of the Lagrangian with respect to the Lagrange multipliers:

$$g_1\left(q_{1,A}\right)-q_{E,1}=0$$
$$g_2\left(q_{2,A}\right)-q_{F,2}=0$$
$$g_3\left(q_{3,A}\right)-q_{G,3}=0$$
$$g_4\left(q_{4,A}\right)-q_{H,4}=0$$
$$g_5\left(q_{5,B}\right)-q_{E,5}=0$$
$$g_6\left(q_{6,B}\right)-q_{F,6}=0$$
$$g_7\left(q_{7,B}\right)-q_{G,7}=0$$
$$g_8\left(q_{8,B}\right)-q_{H,8}=0$$
$$g_9\left(q_{9,C}\right)-q_{E,9}=0$$
$$g_{10}\left(q_{10,C}\right)-q_{F,10}=0$$
$$g_{11}\left(q_{11,C}\right)-q_{G,11}=0$$
$$g_{12}\left(q_{12,C}\right)-q_{H,12}=0$$
$$g_{13}\left(q_{13,D}\right)-q_{E,13}=0$$
$$g_{14}\left(q_{14,D}\right)-q_{F,14}=0$$
$$g_{15}\left(q_{15,D}\right)-q_{G,15}=0$$
$$g_{16}\left(q_{16,D}\right)-q_{H,16}=0$$

4.1.12.5 Eight Balance Equations

The balance equations are the derivative of the Lagrangian with respect to the Lagrange multipliers:

$$q_{A,17}-q_{1,A}-q_{2,A}-q_{3,A}-q_{4,A}=0$$
$$q_{B,18}-q_{5,B}-q_{6,B}-q_{7,B}-q_{8,B}=0$$
$$q_{C,19}-q_{9,C}-q_{10,C}-q_{11,C}-q_{12,C}=0$$

$$q_{D,20} - q_{13,D} - q_{14,D} - q_{15,D} - q_{16,D} = 0$$
$$q_{E,1} + q_{E,5} + q_{E,9} + q_{E,13} - q_{21,E} = 0$$
$$q_{F,2} + q_{F,6} + q_{F,10} + q_{F,14} - q_{22,F} = 0$$
$$q_{G,3} + q_{G,7} + q_{G,11} + q_{G,15} - q_{23,G} = 0$$
$$q_{H,4} + q_{H,8} + q_{H,12} + q_{H,16} - q_{24,H} = 0$$

There are:

- 4 demand equations (in complementarity form)
- 4 supply equations (in complementarity form)
- 16 pairs of equations for each transportation node, one for each input and output (in complementarity form)
- 16 input-output equations (in equality form)
- 8 hub balance equations (in equality form)

This formulation represents 64 equations in the 64 unknowns.

There are 40 complementarity or “perpendicularity” equations and 24 equality equations. That means you have to solve 64 nonlinear inequalities and unknowns if you are to use a direct complementarity type of approach in its complete form. It is a very tall order to put all these complementarity equations together correctly without omission or addition, to actually program them for delivery to a “solver” (generally a GAMS, AMPL, or equivalent “solver”), to write command line code to actually deliver them to the “solver,” to program all the requisite data in myriad individual command line data entry commands, to deliver that data via those myriad individual command lines to the “solver,” to execute the GAMS monolithic “solver” algorithm, to discern whether it is formulated correctly and has worked correctly, to extract the results from the “solver” in an organized fashion, and to report the results. As we shall prove, there is no reason to do so. Our example is a small transportation problem, and one already faces an arduous task with complementarity, extremely so.

It is worthwhile to summarize the “unknowns” being solved for in these equations. There are

- 40 flowing quantities (quantities associated with the links in the network).
- 8 Lagrange multipliers on the node balance constraints (prices associated with the hubs in the network)
- 16 Lagrange multipliers corresponding to the input-output equations.

That gives a total of 64 unknowns. We defer to Section 4.8 the state of the art methods used to solve complementarity questions. Gabriel op. cit. and the literature tell us they are nonlinear

Newton's methods, and they require a Jacobean containing every one of the equations with respect to every one of the unknowns. In this example, a Jacobean solution would require calculating and refreshing an extremely sparse matrix that is 64 by 64 and calculating its inverse on every Newton iteration. (Even though the Jacobean matrix is extremely sparse, its inverse is not. The inverse is far from trivial. And we should note that one can solve a 64 by 64 linear system without actually calculating the inverse per se. Notwithstanding, the number of calculations is still on the same order. When we say "calculating the inverse of the Jacobean," it is acceptable to say solving a linear system involving the Jacobean. Macht nichts; both are ponderous.) Presaging some of the results in this and the next section, this is excessively and unnecessarily large, very much so. We will return to the solution algorithm issues below after making one analytical simplification in this general problem.

One of the most important points to note is the following. For even this simple, highly aggregated model of world gas transportation in Figure 1, if one chooses to use complementarity, he de facto makes **zero attempt and gives zero consideration to any analytic solution to any optimization problem or to any complementarity equation(s)**. In particular, complementarians make zero attempt to analytically solve any of the foregoing 64 equations, and there is no evidence to suggest that they have considered the possibility of doing so. That has been a watchword of complementarity—zero analytic simplification or solution. Head for the computer! Once complementarians have developed their complementarity types of equations, they have been so impatient to get them into numerical solvers that they have overlooked analytical solutions or substitution it seems. That is a glaring problem for the method. **As we shall demonstrate, one can solve all the transportation complementarity equations analytically** and thereby eliminate the need for complementarity solution methods altogether. This saves a huge amount of computer resource, algorithmic complexity, equation writing and programming, labor intensity, propensity for error, and any need to put equations into a "solver." The need for a solver evaporates, the complementarity equations for transportation being amenable to closed-form analytical solution.

The next section is going to simplify the foregoing fully general, fully detailed complementarity formulation by making one useful analytical solution of the input-output function. In fact, this analytic substitution is going to cut the complementarity problem size in half. The substitution emanates from the same "trick" Koopmans and Hitchcock use for linear programming in order to minimize the size of the formulation (which even after such simplification still remains much larger than needed). We will see that even the minimal size complementarity problem is still over four times as large as network microeconomic equilibrium, and many times more complex than needed. Even when it simplifies to linear programming, it is simply too ponderous to represent a transportation matrix of the size that EIA needs to represent, the size we have routinely implemented for world LNG and pipe.

4.2 Substitution of Input-output Relationships Reduces the Size and Complexity of the Complementarity Problem

It is possible to simplify the foregoing formulation; this section illustrates the simplification and derives the key results. To make the simplification, we repeat the same first seven steps in precisely the same way as we did in Section 4.1. However, in Step 8, we make some critical alterations and simplifications. In particular, we substitute the input-output relationships to reduce the number of quantity variables for which we must solve directly. This is a "trick" that has been used in linear programming, and we will deploy it here as well. In particular, the input quantities to the transportation processes are related to the output quantities by the input-output function, and we simply substitute the input-output function.

4.2.1 Step 8: Balance Inputs and Outputs at Every Hub (Outputs Minus Inputs = 0)

This set of equations ensures that the inflow into a node is exactly equal to the output. This is the condition that there can be no excess inflow and no excess outflow to each of the eight hubs. These equations are organized so that the outflows are positive in sign and the inflows are negative in sign, i.e., all equations represent outflows minus inflows.

$$
\begin{gathered}
q_{A,17} - q_{1,A} - q_{2,A} - q_{3,A} - q_{4,A} = 0 \\
q_{B,18} - q_{5,B} - q_{6,B} - q_{7,B} - q_{8,B} = 0 \\
q_{C,19} - q_{9,C} - q_{10,C} - q_{11,C} - q_{12,C} = 0 \\
q_{D,20} - q_{13,D} - q_{14,D} - q_{15,D} - q_{16,D} = 0 \\
q_{E,1} + q_{E,5} + q_{E,9} + q_{E,13} - q_{21,E} = 0 \\
q_{F,2} + q_{F,6} + q_{F,10} + q_{F,14} - q_{22,F} = 0 \\
q_{G,3} + q_{G,7} + q_{G,11} + q_{G,15} - q_{23,G} = 0 \\
q_{H,4} + q_{H,8} + q_{H,12} + q_{H,16} - q_{24,H} = 0
\end{gathered}
$$

We have sixteen input-output relationships, which we can substitute into these eight balance equations. We substitute them directly into the quantity balances and write the amended quantity balance equations at the hubs as follows:

$$
\begin{gathered}
q_{A,17} - q_{1,A} - q_{2,A} - q_{3,A} - q_{4,A} = 0 \\
q_{B,18} - q_{5,B} - q_{6,B} - q_{7,B} - q_{8,B} = 0 \\
q_{C,19} - q_{9,C} - q_{10,C} - q_{11,C} - q_{12,C} = 0 \\
q_{D,20} - q_{13,D} - q_{14,D} - q_{15,D} - q_{16,D} = 0 \\
g_1\left(q_{1,A}\right) + g_5\left(q_{5,B}\right) + g_9\left(q_{9,C}\right) + g_{13}\left(q_{13,D}\right) - q_{21,E} = 0 \\
g_2\left(q_{2,A}\right) + g_6\left(q_{6,B}\right) + g_{10}\left(q_{10,C}\right) + g_{14}\left(q_{14,D}\right) - q_{22,F} = 0 \\
g_3\left(q_{3,A}\right) + g_7\left(q_{7,B}\right) + g_{11}\left(q_{11,C}\right) + g_{15}\left(q_{15,D}\right) - q_{23,G} = 0 \\
g_4\left(q_{4,A}\right) + g_8\left(q_{8,B}\right) + g_{12}\left(q_{12,C}\right) + g_{16}\left(q_{16,D}\right) - q_{24,H} = 0
\end{gathered}
$$

Notice that we have eliminated via substitution the quantity flows that correspond to the input links and replaced them with the input-output functions. Such substitution reduces the number of variables.

4.2.2 Step 9: Enumerate the Nonnegativity Constraints Everywhere

Every quantity on every link must have a non-negativity constraint.

$$
\begin{array}{llll}
q_{1,A} \ge 0 & q_{2,A} \ge 0 & q_{3,A} \ge 0 & q_{4,A} \ge 0 \\
q_{5,B} \ge 0 & q_{6,B} \ge 0 & q_{7,B} \ge 0 & q_{8,B} \ge 0 \\
q_{9,C} \ge 0 & q_{10,C} \ge 0 & q_{11,C} \ge 0 & q_{12,C} \ge 0 \\
q_{13,D} \ge 0 & q_{14,D} \ge 0 & q_{15,D} \ge 0 & q_{16,D} \ge 0 \\
q_{A,17} \ge 0 & q_{B,18} \ge 0 & q_{C,10} \ge 0 & q_{D,20} \ge 0 \\
q_{21,E} \ge 0 & q_{22,F} \ge 0 & q_{23,G} \ge 0 & q_{24,H} \ge 0
\end{array}
$$

4.2.3 Step 10: Write the Lagrangian

The Lagrangian has a simpler form having omitted the sixteen variables that are the inputs to the transportation nodes.

$$
\begin{aligned}
L = & -CS_{17}\left(q_{A,17}\right) - CS_{18}\left(q_{B,18}\right) - CS_{19}\left(q_{C,19}\right) - CS_{20}\left(q_{D,20}\right) \\
& + VOC_1 q_{1,A} + VOC_2 q_{2,A} + VOC_3 q_{3,A} + VOC_4 q_{4,A} + VOC_5 q_{5,B} + VOC_6 q_{6,B} + VOC_7 q_{7,B} + VOC_8 q_{8,B} \\
& + VOC_9 q_{9,C} + VOC_{10} q_{10,C} + VOC_{11} q_{11,C} + VOC_{12} q_{12,C} + VOC_{13} q_{13,D} + VOC_{14} q_{14,D} + VOC_{15} q_{15,D} + VOC_{16} q_{16,D} \\
& + TC_{21}\left(q_{21,E}\right) + TC_{22}\left(q_{22,F}\right) + TC_{23}\left(q_{23,G}\right) + TC_{24}\left(q_{24,H}\right) \\
& + \lambda_{A,17}\left[q_{A,17} - q_{1,A} - q_{2,A} - q_{3,A} - q_{4,A}\right] + \lambda_{B,18}\left[q_{B,18} - q_{5,B} - q_{6,B} - q_{7,B} - q_{8,B}\right] \\
& + \lambda_{C,19}\left[q_{C,19} - q_{9,C} - q_{10,C} - q_{11,C} - q_{12,C}\right] + \lambda_{D,20}\left[q_{D,20} - q_{13,D} - q_{14,D} - q_{15,D} - q_{16,D}\right] \\
& + \lambda_{21,E}\left[g_1\left(q_{1,A}\right) + g_5\left(q_{5,B}\right) + g_9\left(q_{9,C}\right) + g_{13}\left(q_{13,D}\right) - q_{21,E}\right] + \lambda_{22,F}\left[g_2\left(q_{2,A}\right) + g_6\left(q_{6,B}\right) + g_{10}\left(q_{10,C}\right) + g_{14}\left(q_{14,D}\right) - q_{22,F}\right] \\
& + \lambda_{23,G}\left[g_3\left(q_{3,A}\right) + g_7\left(q_{7,B}\right) + g_{11}\left(q_{11,C}\right) + g_{15}\left(q_{15,D}\right) - q_{23,G}\right] + \lambda_{24,H}\left[g_4\left(q_{4,A}\right) + g_8\left(q_{8,B}\right) + g_{12}\left(q_{12,C}\right) + g_{16}\left(q_{16,D}\right) - q_{24,H}\right] \\
& - \mu_{1,A} q_{1,A} - \mu_{2,A} q_{2,A} - \mu_{3,A} q_{3,A} - \mu_{4,A} q_{4,A} - \mu_{5,B} q_{5,B} - \mu_{6,B} q_{6,B} - \mu_{7,B} q_{7,B} - \mu_{8,B} q_{8,B} \\
& - \mu_{9,C} q_{9,C} - \mu_{10,C} q_{10,C} - \mu_{11,C} q_{11,C} - \mu_{12,C} q_{12,C} - \mu_{13,D} q_{13,D} - \mu_{14,D} q_{14,D} - \mu_{15,D} q_{15,D} - \mu_{16,D} q_{16,D} \\
& - \mu_{A,17} q_{A,17} - \mu_{B,18} q_{B,18} - \mu_{C,19} q_{C,19} - \mu_{D,20} q_{D,20} - \mu_{21,E} q_{21,E} - \mu_{22,F} q_{22,F} - \mu_{23,G} q_{23,G} - \mu_{24,H} q_{24,H}
\end{aligned}
$$

There are now only 8 Lagrange multipliers for 8 hub constraints. There are 4+4+16 quantities instead of the former 4+4+16+16. We have reduced by 16 the number of Lagrange multipliers and reduced by 16 the number quantities. This reduces the total to 32. In particular, there are now 24 unknown quantities and 8 unknown Lagrange multipliers (the latter being interpreted as prices). As we will see, this is still way too many, but it is as small as we can go with complementarity.

This section begins to presage just how much leverage you get when you solve Kuhn-Tucker conditions **analytically** (by substitution) rather than defaulting unthinkingly to numerical methods. With the simplest of substitutions—the input-output relationships—the size of the problem has been cut in half. We shall see that the size of the problem is cut by another factor of four with algebraic solution of the complementarity equations. Algebraic substitution is the prudent way to solve the equations, not numerics. One should never approach complementarity numerically when it is so amenable to analytic substitution and solution.

4.2.4 Step 11: Compute the Twenty-Four First Derivatives, Which Are the Kuhn-Tucker Conditions

Using this formulation, we calculate the requisite first derivatives of the Lagrangian.

4.2.4.1 Derivatives of Demand Equations (Four)

$$\frac{\partial L}{\partial q_{A,17}} = -p_{17}\left(q_{A,17}\right) + \lambda_{A,17} - \mu_{A,17} = 0$$
$$\frac{\partial L}{\partial q_{B,18}} = -p_{18}\left(q_{B,18}\right) + \lambda_{B,18} - \mu_{B,18} = 0$$
$$\frac{\partial L}{\partial q_{C,19}} = -p_{19}\left(q_{C,19}\right) + \lambda_{C,19} - \mu_{C,19} = 0$$
$$\frac{\partial L}{\partial q_{D,20}} = -p_{20}\left(q_{D,20}\right) + \lambda_{D,20} - \mu_{D,20} = 0$$

4.2.4.2 Derivatives of Supply Equations (Four)

$$\frac{\partial L}{\partial q_{21,E}} = MC_{21}\left(q_{21,E}\right) - \lambda_{21,E} - \mu_{21,E} = 0$$
$$\frac{\partial L}{\partial q_{22,F}} = MC_{22}\left(q_{22,F}\right) - \lambda_{22,F} - \mu_{22,F} = 0$$
$$\frac{\partial L}{\partial q_{23,G}} = MC_{23}\left(q_{23,G}\right) - \lambda_{23,G} - \mu_{23,G} = 0$$
$$\frac{\partial L}{\partial q_{24,H}} = MC_{24}\left(q_{24,H}\right) - \lambda_{24,H} - \mu_{24,H} = 0$$

4.2.4.3 Derivatives of Outputs from Transportation Activities (Sixteen)

$$\frac{\partial L}{\partial q_{1,A}} = VOC_1 + \lambda_{21,E} g_1'\left(q_{1,A}\right) - \lambda_{A,17} - \mu_{1,A} = 0$$
$$\frac{\partial L}{\partial q_{2,A}} = VOC_2 + \lambda_{22,F} g_2'\left(q_{2,A}\right) - \lambda_{A,17} - \mu_{2,A} = 0$$
$$\frac{\partial L}{\partial q_{3,A}} = VOC_3 + \lambda_{23,G} g_3'\left(q_{3,A}\right) - \lambda_{A,17} - \mu_{3,A} = 0$$
$$\frac{\partial L}{\partial q_{4,A}} = VOC_4 + \lambda_{24,H} g_4'\left(q_{4,A}\right) - \lambda_{A,17} - \mu_{4,A} = 0$$
$$\frac{\partial L}{\partial q_{5,B}} = VOC_5 + \lambda_{21,E} g_5'\left(q_{5,B}\right) - \lambda_{B,18} - \mu_{5,B} = 0$$
$$\frac{\partial L}{\partial q_{6,B}} = VOC_6 + \lambda_{22,F} g_6'\left(q_{6,B}\right) - \lambda_{B,18} - \mu_{6,B} = 0$$

$$\frac{\partial L}{\partial q_{7,B}} = VOC_7 + \lambda_{23,G} g_7 '(q_{7B}) - \lambda_{B,18} - \mu_{7,B} = 0$$
$$\frac{\partial L}{\partial q_{8,B}} = VOC_8 + \lambda_{24,H} g_8 '(q_{8,B}) - \lambda_{B,18} - \mu_{8,B} = 0$$
$$\frac{\partial L}{\partial q_{9,C}} = VOC_9 + \lambda_{21,E} g_9 '(q_{9,C}) - \lambda_{C,19} - \mu_{9,C} = 0$$
$$\frac{\partial L}{\partial q_{10,C}} = VOC_{10} + \lambda_{22,F} g_{10} '(q_{10,C}) - \lambda_{C,19} - \mu_{10,C} = 0$$
$$\frac{\partial L}{\partial q_{11,C}} = VOC_{11} + \lambda_{23,G} g_{11} '(q_{11,C}) - \lambda_{C,19} - \mu_{11,C} = 0$$
$$\frac{\partial L}{\partial q_{12,C}} = VOC_{12} + \lambda_{24,H} g_{12} '(q_{12,C}) - \lambda_{C,19} - \mu_{12,C} = 0$$
$$\frac{\partial L}{\partial q_{13,D}} = VOC_{13} + \lambda_{21,E} g_{13} '(q_{13,D}) - \lambda_{D,20} - \mu_{13,D} = 0$$
$$\frac{\partial L}{\partial q_{14,D}} = VOC_{14} + \lambda_{22,F} g_{14} '(q_{14,D}) - \lambda_{D,20} - \mu_{14,D} = 0$$
$$\frac{\partial L}{\partial q_{15,D}} = VOC_{15} + \lambda_{23,G} g_{15} '(q_{15,D}) - \lambda_{D,20} - \mu_{15,D} = 0$$
$$\frac{\partial L}{\partial q_{16,D}} = VOC_{16} + \lambda_{24,H} g_{16} '(q_{16,D}) - \lambda_{D,20} - \mu_{16,D} = 0$$

4.2.4.4 Hub Balance Equations (Eight)

$$\frac{\partial L}{\partial \lambda_{A,17}} = q_{A,17} - q_{1,A} - q_{2,A} - q_{3,A} - q_{4,A} = 0$$
$$\frac{\partial L}{\partial \lambda_{B,18}} = q_{B,18} - q_{5,B} - q_{6,B} - q_{7,B} - q_{8,B} = 0$$
$$\frac{\partial L}{\partial \lambda_{C,19}} = q_{C,19} - q_{9,C} - q_{10,C} - q_{11,C} - q_{12,C} = 0$$
$$\frac{\partial L}{\partial \lambda_{D,20}} = q_{D,20} - q_{13,D} - q_{14,D} - q_{15,D} - q_{16,D} = 0$$
$$\frac{\partial L}{\partial \lambda_{21,E}} = g_1(q_{1,A}) + g_5(q_{5,B}) + g_9(q_{9,C}) + g_{13}(q_{13,D}) - q_{21,E} = 0$$
$$\frac{\partial L}{\partial \lambda_{22,F}} = g_2(q_{2,A}) + g_6(q_{6,B}) + g_{10}(q_{10,C}) + g_{14}(q_{14,D}) - q_{22,F} = 0$$
$$\frac{\partial L}{\partial \lambda_{23,G}} = g_3(q_{3,A}) + g_7(q_{7,B}) + g_{11}(q_{11,C}) + g_{15}(q_{15,D}) - q_{23,G} = 0$$
$$\frac{\partial L}{\partial \lambda_{24,H}} = g_4(q_{4,A}) + g_8(q_{8,B}) + g_{12}(q_{12,C}) + g_{16}(q_{16,D}) - q_{24,H} = 0$$

4.2.4.5 Inequality Kuhn-Tucker Conditions

$$\begin{array}{lll}
\mu_{1,A} q_{1,A} = 0 & \mu_{1,A} \ge 0 & q_{1,A} \ge 0 \\
\mu_{2,A} q_{2,A} = 0 & \mu_{2,A} \ge 0 & q_{2,A} \ge 0 \\
\mu_{3,A} q_{3,A} = 0 & \mu_{3,A} \ge 0 & q_{3,A} \ge 0 \\
\mu_{4,A} q_{4,A} = 0 & \mu_{4,A} \ge 0 & q_{4,A} \ge 0 \\
\mu_{5,B} q_{5,B} = 0 & \mu_{5,B} \ge 0 & q_{5,B} \ge 0 \\
\mu_{6,B} q_{6,B} = 0 & \mu_{6,B} \ge 0 & q_{6,B} \ge 0 \\
\mu_{7,B} q_{7,B} = 0 & \mu_{7,B} \ge 0 & q_{7,B} \ge 0 \\
\mu_{8,B} q_{8,B} = 0 & \mu_{8,B} \ge 0 & q_{8,B} \ge 0 \\
\mu_{9,C} q_{9,C} = 0 & \mu_{9,C} \ge 0 & q_{9,C} \ge 0 \\
\mu_{10,C} q_{10,C} = 0 & \mu_{10,C} \ge 0 & q_{10,C} \ge 0 \\
\mu_{11,C} q_{11,C} = 0 & \mu_{11,C} \ge 0 & q_{11,C} \ge 0 \\
\mu_{12,C} q_{12,C} = 0 & \mu_{12,C} \ge 0 & q_{12,C} \ge 0 \\
\mu_{13,D} q_{13,D} = 0 & \mu_{13,D} \ge 0 & q_{13,D} \ge 0 \\
\mu_{14,D} q_{14,D} = 0 & \mu_{14,D} \ge 0 & q_{14,D} \ge 0 \\
\mu_{15,D} q_{15,D} = 0 & \mu_{15,D} \ge 0 & q_{15,D} \ge 0 \\
\mu_{16,D} q_{16,D} = 0 & \mu_{16,D} \ge 0 & q_{16,D} \ge 0 \\
\mu_{A,17} q_{A,17} = 0 & \mu_{A,17} \ge 0 & q_{A,17} \ge 0 \\
\mu_{B,18} q_{B,18} = 0 & \mu_{B,18} \ge 0 & q_{B,18} \ge 0 \\
\mu_{C,19} q_{C,19} = 0 & \mu_{C,19} \ge 0 & q_{C,19} \ge 0 \\
\mu_{D,20} q_{D,20} = 0 & \mu_{D,20} \ge 0 & q_{D,20} \ge 0 \\
\mu_{21,E} q_{21,E} = 0 & \mu_{21,E} \ge 0 & q_{21,E} \ge 0 \\
\mu_{22,F} q_{22,F} = 0 & \mu_{22,F} \ge 0 & q_{22,F} \ge 0 \\
\mu_{23,G} q_{23,G} = 0 & \mu_{23,G} \ge 0 & q_{23,G} \ge 0 \\
\mu_{24,H} q_{24,H} = 0 & \mu_{24,H} \ge 0 & q_{24,H} \ge 0
\end{array}$$

Even after the simplification the input-output substitution allows, this is still a formidable set of Kuhn-Tucker conditions.

4.2.5 Step 12: Write the Kuhn-Tucker Conditions Using These Derivatives

This section writes the Kuhn-Tucker conditions

4.2.5.1 Derivatives of Demand Equations (Four)

$$\begin{aligned}
-p_{17}\left(q_{A,17}\right) + \lambda_{A,17} &= \mu_{A,17} \\
-p_{18}\left(q_{B,18}\right) + \lambda_{B,18} &= \mu_{B,18} \\
-p_{19}\left(q_{C,19}\right) + \lambda_{C,19} &= \mu_{C,19} \\
-p_{20}\left(q_{D,20}\right) + \lambda_{D,20} &= \mu_{D,20}
\end{aligned}$$

4.2.5.2 Derivatives of Supply Equations (Four)

$$MC_{21}\left(q_{21,E}\right)-\lambda_{21,E}=\mu_{21,E}$$
$$MC_{22}\left(q_{22,F}\right)-\lambda_{22,F}=\mu_{22,F}$$
$$MC_{23}\left(q_{23,G}\right)-\lambda_{23,G}=\mu_{23,G}$$
$$MC_{24}\left(q_{24,H}\right)-\lambda_{24,H}=\mu_{24,H}$$

4.2.5.3 Derivatives of Outputs from Transportation Activities (Sixteen)

$$VOC_1+\lambda_{21,E}g_1{}'\left(q_{1,A}\right)-\lambda_{A,17}=\mu_{1,A}$$
$$VOC_2+\lambda_{22,F}g_2{}'\left(q_{2,A}\right)-\lambda_{A,17}=\mu_{2,A}$$
$$VOC_3+\lambda_{23,G}g_3{}'\left(q_{3,A}\right)-\lambda_{A,17}=\mu_{3,A}$$
$$VOC_4+\lambda_{24,H}g_4{}'\left(q_{4,A}\right)-\lambda_{A,17}=\mu_{4,A}$$
$$VOC_5+\lambda_{21,E}g_5{}'\left(q_{5,B}\right)-\lambda_{B,18}=\mu_{5,B}$$
$$VOC_6+\lambda_{22,F}g_6{}'\left(q_{6,B}\right)-\lambda_{B,18}=\mu_{6,B}$$
$$VOC_7+\lambda_{23,G}g_7{}'\left(q_{7B}\right)-\lambda_{B,18}=\mu_{7,B}$$
$$VOC_8+\lambda_{24,H}g_8{}'\left(q_{8,B}\right)-\lambda_{B,18}=\mu_{8,B}$$
$$VOC_9+\lambda_{21,E}g_9{}'\left(q_{9,C}\right)-\lambda_{C,19}=\mu_{9,C}$$
$$VOC_{10}+\lambda_{22,F}g_{10}{}'\left(q_{10,C}\right)-\lambda_{C,19}=\mu_{10,C}$$
$$VOC_{11}+\lambda_{23,G}g_{11}{}'\left(q_{11,C}\right)-\lambda_{C,19}=\mu_{11,C}$$
$$VOC_{12}+\lambda_{24,H}g_{12}{}'\left(q_{12,C}\right)-\lambda_{C,19}=\mu_{12,C}$$
$$VOC_{13}+\lambda_{21,E}g_{13}{}'\left(q_{13,D}\right)-\lambda_{D,20}=\mu_{13,D}$$
$$VOC_{14}+\lambda_{22,F}g_{14}{}'\left(q_{14,D}\right)-\lambda_{D,20}=\mu_{14,D}$$
$$VOC_{15}+\lambda_{23,G}g_{15}{}'\left(q_{15,D}\right)-\lambda_{D,20}=\mu_{15,D}$$
$$VOC_{16}+\lambda_{24,H}g_{16}{}'\left(q_{16,D}\right)-\lambda_{D,20}=\mu_{16,D}$$

4.2.5.4 Hub Balance Equations (Eight)

$$q_{A,17}-q_{1,A}-q_{2,A}-q_{3,A}-q_{4,A}=0$$
$$q_{B,18}-q_{5,B}-q_{6,B}-q_{7,B}-q_{8,B}=0$$
$$q_{C,19}-q_{9,C}-q_{10,C}-q_{11,C}-q_{12,C}=0$$
$$q_{D,20}-q_{13,D}-q_{14,D}-q_{15,D}-q_{16,D}=0$$

$$g_1(q_{1,A}) + g_5(q_{5,B}) + g_9(q_{9,C}) + g_{13}(q_{13,D}) - q_{21,E} = 0$$
$$g_2(q_{2,A}) + g_6(q_{6,B}) + g_{10}(q_{10,C}) + g_{14}(q_{14,D}) - q_{22,F} = 0$$
$$g_3(q_{3,A}) + g_7(q_{7,B}) + g_{11}(q_{11,C}) + g_{15}(q_{15,D}) - q_{23,G} = 0$$
$$g_4(q_{4,A}) + g_8(q_{8,B}) + g_{12}(q_{12,C}) + g_{16}(q_{16,D}) - q_{24,H} = 0$$

4.2.5.5 Inequality Kuhn-Tucker Conditions

$$\begin{array}{lll}
\mu_{1,A}q_{1,A} = 0 & \mu_{1,A} \geq 0 & q_{1,A} \geq 0 \\
\mu_{2,A}q_{2,A} = 0 & \mu_{2,A} \geq 0 & q_{2,A} \geq 0 \\
\mu_{3,A}q_{3,A} = 0 & \mu_{3,A} \geq 0 & q_{3,A} \geq 0 \\
\mu_{4,A}q_{4,A} = 0 & \mu_{4,A} \geq 0 & q_{4,A} \geq 0 \\
\mu_{5,B}q_{5,B} = 0 & \mu_{5,B} \geq 0 & q_{5,B} \geq 0 \\
\mu_{6,B}q_{6,B} = 0 & \mu_{6,B} \geq 0 & q_{6,B} \geq 0 \\
\mu_{7,B}q_{7,B} = 0 & \mu_{7,B} \geq 0 & q_{7,B} \geq 0 \\
\mu_{8,B}q_{8,B} = 0 & \mu_{8,B} \geq 0 & q_{8,B} \geq 0 \\
\mu_{9,C}q_{9,C} = 0 & \mu_{9,C} \geq 0 & q_{9,C} \geq 0 \\
\mu_{10,C}q_{10,C} = 0 & \mu_{10,C} \geq 0 & q_{10,C} \geq 0 \\
\mu_{11,C}q_{11,C} = 0 & \mu_{11,C} \geq 0 & q_{11,C} \geq 0 \\
\mu_{12,C}q_{12,C} = 0 & \mu_{12,C} \geq 0 & q_{12,C} \geq 0 \\
\mu_{13,D}q_{13,D} = 0 & \mu_{13,D} \geq 0 & q_{13,D} \geq 0 \\
\mu_{14,D}q_{14,D} = 0 & \mu_{14,D} \geq 0 & q_{14,D} \geq 0 \\
\mu_{15,D}q_{15,D} = 0 & \mu_{15,D} \geq 0 & q_{15,D} \geq 0 \\
\mu_{16,D}q_{16,D} = 0 & \mu_{16,D} \geq 0 & q_{16,D} \geq 0 \\
\mu_{A,17}q_{A,17} = 0 & \mu_{A,17} \geq 0 & q_{A,17} \geq 0 \\
\mu_{B,18}q_{B,18} = 0 & \mu_{B,18} \geq 0 & q_{B,18} \geq 0 \\
\mu_{C,19}q_{C,19} = 0 & \mu_{C,19} \geq 0 & q_{C,19} \geq 0 \\
\mu_{D,20}q_{D,20} = 0 & \mu_{D,20} \geq 0 & q_{D,20} \geq 0 \\
\mu_{21,E}q_{21,E} = 0 & \mu_{21,E} \geq 0 & q_{21,E} \geq 0 \\
\mu_{22,F}q_{22,F} = 0 & \mu_{22,F} \geq 0 & q_{22,F} \geq 0 \\
\mu_{23,G}q_{23,G} = 0 & \mu_{23,G} \geq 0 & q_{23,G} \geq 0 \\
\mu_{24,H}q_{24,H} = 0 & \mu_{24,H} \geq 0 & q_{24,H} \geq 0
\end{array}$$

4.2.6 Step 13: Write These Kuhn-Tucker Conditions in Complementarity Form

This section writes the Kuhn-Tucker conditions in complementarity form, deriving them from the foregoing derivatives by eliminating the Lagrange multipliers on the inequality constraints.

4.2.6.1 Four Demand Equations (4 Total)

$$\begin{array}{lll}
\left[\lambda_{A,17} - p_{17}(q_{A,17})\right] q_{A,17} = 0 & \left[\lambda_{A,17} - p_{17}(q_{A,17})\right] \geq 0 & q_{A,17} \geq 0 \\
\left[\lambda_{B,18} - p_{18}(q_{B,18})\right] q_{B,18} = 0 & \left[\lambda_{B,18} - p_{18}(q_{B,18})\right] \geq 0 & q_{B,18} \geq 0 \\
\left[\lambda_{C,19} - p_{19}(q_{C,19})\right] q_{C,19} = 0 & \left[\lambda_{C,19} - p_{19}(q_{C,19})\right] \geq 0 & q_{C,19} \geq 0
\end{array}$$

$$\left[\lambda_{D,20}-p_{20}\left(q_{D,20}\right)\right]q_{D,20}=0 \qquad \left[\lambda_{D,20}-p_{20}\left(q_{D,20}\right)\right]\geq 0 \qquad q_{D,20}\geq 0$$

4.2.6.2 Four Supply Equations (4 Total)

$$\left[MC_{21}\left(q_{21,E}\right)-\lambda_{21,E}\right]q_{21,E}=0 \qquad \left[MC_{21}\left(q_{21,E}\right)-\lambda_{21,E}\right]\geq 0 \qquad q_{21,E}\geq 0$$
$$\left[MC_{22}\left(q_{22,F}\right)-\lambda_{22,F}\right]q_{22,F}=0 \qquad \left[MC_{22}\left(q_{22,F}\right)-\lambda_{22,F}\right]\geq 0 \qquad q_{22,F}\geq 0$$
$$\left[MC_{23}\left(q_{23,G}\right)-\lambda_{23,G}\right]q_{23,G}=0 \qquad \left[MC_{23}\left(q_{23,G}\right)-\lambda_{23,G}\right]\geq 0 \qquad q_{23,G}\geq 0$$
$$\left[MC_{24}\left(q_{24,H}\right)-\lambda_{24,H}\right]q_{24,H}=0 \qquad \left[MC_{24}\left(q_{24,H}\right)-\lambda_{24,H}\right]\geq 0 \qquad q_{24,H}\geq 0$$

4.2.6.3 Sixteen Transportation Output Relationships (16 Total)

$$\left[VOC_1+\lambda_{21,E}g_1'\left(q_{1,A}\right)-\lambda_{A,17}\right]q_{1,A}=0 \qquad \left[VOC_1+\lambda_{21,E}g_1'\left(q_{1,A}\right)-\lambda_{A,17}\right]\geq 0 \qquad q_{1,A}\geq 0$$
$$\left[VOC_2+\lambda_{22,F}g_2'\left(q_{2,A}\right)-\lambda_{A,17}\right]q_{2,A}=0 \qquad \left[VOC_2+\lambda_{22,F}g_2'\left(q_{2,A}\right)-\lambda_{A,17}\right]\geq 0 \qquad q_{2,A}\geq 0$$
$$\left[VOC_3+\lambda_{23,G}g_3'\left(q_{3,A}\right)-\lambda_{A,17}\right]q_{3,A}=0 \qquad \left[VOC_3+\lambda_{23,G}g_3'\left(q_{3,A}\right)-\lambda_{A,17}\right]\geq 0 \qquad q_{3,A}\geq 0$$
$$\left[VOC_4+\lambda_{24,H}g_4'\left(q_{4,A}\right)-\lambda_{A,17}\right]q_{4,A}=0 \qquad \left[VOC_4+\lambda_{24,H}g_4'\left(q_{4,A}\right)-\lambda_{A,17}\right]\geq 0 \qquad q_{4,A}\geq 0$$
$$\left[VOC_5+\lambda_{21,E}g_5'\left(q_{5,B}\right)-\lambda_{B,18}\right]q_{5,B}=0 \qquad \left[VOC_5+\lambda_{21,E}g_5'\left(q_{5,B}\right)-\lambda_{B,18}\right]\geq 0 \qquad q_{5,B}\geq 0$$
$$\left[VOC_6+\lambda_{22,F}g_6'\left(q_{6,B}\right)-\lambda_{B,18}\right]q_{6,B}=0 \qquad \left[VOC_6+\lambda_{22,F}g_6'\left(q_{6,B}\right)-\lambda_{B,18}\right]\geq 0 \qquad q_{6,B}\geq 0$$
$$\left[VOC_7+\lambda_{23,G}g_7'\left(q_{7B}\right)-\lambda_{B,18}\right]q_{7B}=0 \qquad \left[VOC_7+\lambda_{23,G}g_7'\left(q_{7B}\right)-\lambda_{B,18}\right]\geq 0 \qquad q_{7B}\geq 0$$
$$\left[VOC_8+\lambda_{24,H}g_8'\left(q_{8,B}\right)-\lambda_{B,18}\right]q_{8,B}=0 \qquad \left[VOC_8+\lambda_{24,H}g_8'\left(q_{8,B}\right)-\lambda_{B,18}\right]\geq 0 \qquad q_{8,B}\geq 0$$
$$-\left[VOC_9+\lambda_{21,E}g_9'\left(q_{9,C}\right)-\lambda_{C,19}\right]q_{9,C}=0 \qquad \left[VOC_9+\lambda_{21,E}g_9'\left(q_{9,C}\right)-\lambda_{C,19}\right]\geq 0 \qquad q_{9,C}\geq 0$$
$$\left[VOC_{10}+\lambda_{22,F}g_{10}'\left(q_{10,C}\right)-\lambda_{C,19}\right]q_{10,C}=0 \qquad \left[VOC_{10}+\lambda_{22,F}g_{10}'\left(q_{10,C}\right)-\lambda_{C,19}\right]\geq 0 \qquad q_{10,C}\geq 0$$
$$\left[VOC_{11}+\lambda_{23,G}g_{11}'\left(q_{11,C}\right)-\lambda_{C,19}\right]q_{11,C}=0 \qquad \left[VOC_{11}+\lambda_{23,G}g_{11}'\left(q_{11,C}\right)-\lambda_{C,19}\right]\geq 0 \qquad q_{11,C}\geq 0$$
$$\left[VOC_{12}+\lambda_{24,H}g_{12}'\left(q_{12,C}\right)-\lambda_{C,19}\right]q_{12,C}=0 \qquad \left[VOC_{12}+\lambda_{24,H}g_{12}'\left(q_{12,C}\right)-\lambda_{C,19}\right]\geq 0 \qquad q_{12,C}\geq 0$$
$$\left[VOC_{13}+\lambda_{21,E}g_{13}'\left(q_{13,D}\right)-\lambda_{D,20}\right]q_{13,D}=0 \qquad \left[VOC_{13}+\lambda_{21,E}g_{13}'\left(q_{13,D}\right)-\lambda_{D,20}\right]\geq 0 \qquad q_{13,D}\geq 0$$
$$\left[VOC_{14}+\lambda_{22,F}g_{14}'\left(q_{14,D}\right)-\lambda_{D,20}\right]q_{14,D}=0 \qquad \left[VOC_{14}+\lambda_{22,F}g_{14}'\left(q_{14,D}\right)-\lambda_{D,20}\right]\geq 0 \qquad q_{14,D}\geq 0$$
$$\left[VOC_{15}+\lambda_{23,G}g_{15}'\left(q_{15,D}\right)-\lambda_{D,20}\right]q_{15,D}=0 \qquad \left[VOC_{15}+\lambda_{23,G}g_{15}'\left(q_{15,D}\right)-\lambda_{D,20}\right]\geq 0 \qquad q_{15,D}\geq 0$$
$$\left[VOC_{16}+\lambda_{24,H}g_{16}'\left(q_{16,D}\right)-\lambda_{D,20}\right]q_{16,D}=0 \qquad \left[VOC_{16}+\lambda_{24,H}g_{16}'\left(q_{16,D}\right)-\lambda_{D,20}\right]\geq 0 \qquad q_{16,D}\geq 0$$

4.2.6.4 Eight Balance Equations

$$q_{A,17}-q_{1,A}-q_{2,A}-q_{3,A}-q_{4,A}=0$$
$$q_{B,18}-q_{5,B}-q_{6,B}-q_{7,B}-q_{8,B}=0$$
$$q_{C,19}-q_{9,C}-q_{10,C}-q_{11,C}-q_{12,C}=0$$

$$q_{D,20} - q_{13,D} - q_{14,D} - q_{15,D} - q_{16,D} = 0$$
$$g_1\left(q_{1,A}\right) + g_5\left(q_{5,B}\right) + g_9\left(q_{9,C}\right) + g_{13}\left(q_{13,D}\right) - q_{21,E} = 0$$
$$g_2\left(q_{2,A}\right) + g_6\left(q_{6,B}\right) + g_{10}\left(q_{10,C}\right) + g_{14}\left(q_{14,D}\right) - q_{22,F} = 0$$
$$g_3\left(q_{3,A}\right) + g_7\left(q_{7,B}\right) + g_{11}\left(q_{11,C}\right) + g_{15}\left(q_{15,D}\right) - q_{23,G} = 0$$
$$g_4\left(q_{4,A}\right) + g_8\left(q_{8,B}\right) + g_{12}\left(q_{12,C}\right) + g_{16}\left(q_{16,D}\right) - q_{24,H} = 0$$

There are eight Lagrange multipliers (which will be prices) and sixteen transportation outputs, 24 unknowns in all. However, there are also four aggregate demands and four aggregate supplies, which are added to the system via the material balance equations at the nodes. This will give a grand total of 32 equations and 32 unknowns.

This is substantially simpler and more intuitive problem than the original problem, which had thirty two more equations and unknowns, exactly twice as many. We have reduced the prior number by half simply by substituting the input-output relationships and solving a simpler problem. Therein lies a strong hint—**analytically substitute whenever and wherever you can**. The foregoing problem is still way too difficult and cumbersome to solve, substantially too much so for EIA's needs. There is no reason to solve a problem this large and complex; the problem still admits of **even more analytical rather than numerical solution** as we shall see in subsequent sections. Furthermore, with this type of equation structure, it is tremendously difficult to manipulate the equations when nodes are added or deleted or problem structure is added. It is a computer architectural challenge of insurmountable proportion, fraught with potential error. And for every node you add or delete, you have to program or redact equations for that node. There is simply no reason to bear this much cost and endure this much potential for error.

4.3 Derive an Embedded Transportation Cost Minimization Problem

It has long been known in transportation-logistic-network problems that there is a transportation cost minimization problem embedded within the monolithic global welfare maximization problem.[23] By unearthing and highlighting this embedded transportation cost minimization problem, we will show definitively that linear programming as practiced by the EIA and others is insufficient to solve economic problems. Linear programming is not enough. People who would assert that linear programming is a complete solution are shown not to be correct; more than linear programming is required. We begin by rewriting the Lagrangian for the monolithic global welfare maximization problem. If we segregate terms, we can identify the embedded transportation cost minimization problem easily. The original Lagrangian can be rearranged as follows:

$$\begin{aligned}
L = & -CS_{17}\left(q_{A,17}\right) - CS_{18}\left(q_{B,18}\right) - CS_{19}\left(q_{C,19}\right) - CS_{20}\left(q_{D,20}\right) \\
& + TC_{21}\left(q_{21,E}\right) + TC_{22}\left(q_{22,F}\right) + TC_{23}\left(q_{23,G}\right) + TC_{24}\left(q_{24,H}\right) \\
& - \mu_{A,17} q_{A,17} - \mu_{B,18} q_{B,18} - \mu_{C,19} q_{C,19} - \mu_{D,20} q_{D,20} - \mu_{21,E} q_{21,E} - \mu_{22,F} q_{22,F} - \mu_{23,G} q_{23,G} - \mu_{24,H} q_{24,H} \\
& --
\end{aligned}$$

[23] This was one of the breakthroughs from Samuelson in 1952.

$$
\begin{aligned}
&+VOC_1 q_{1,A} + VOC_2 q_{2,A} + VOC_3 q_{3,A} + VOC_4 q_{4,A} + VOC_5 q_{5,B} + VOC_6 q_{6,B} + VOC_7 q_{7,B} + VOC_8 q_{8,B} \\
&+VOC_9 q_{9,C} + VOC_{10} q_{10,C} + VOC_{11} q_{11,C} + VOC_{12} q_{12,C} + VOC_{13} q_{13,D} + VOC_{14} q_{14,D} + VOC_{15} q_{15,D} + VOC_{16} q_{16,D} \\
&+\lambda_{A,17}\left[q_{A,17} - q_{1,A} - q_{2,A} - q_{3,A} - q_{4,A}\right] + \lambda_{B,18}\left[q_{B,18} - q_{5,B} - q_{6,B} - q_{7,B} - q_{8,B}\right] \\
&+\lambda_{C,19}\left[q_{C,19} - q_{9,C} - q_{10,C} - q_{11,C} - q_{12,C}\right] + \lambda_{D,20}\left[q_{D,20} - q_{13,D} - q_{14,D} - q_{15,D} - q_{16,D}\right] \\
&+\lambda_{21,E}\left[g_1\left(q_{1,A}\right) + g_5\left(q_{5,B}\right) + g_9\left(q_{9,C}\right) + g_{13}\left(q_{13,D}\right) - q_{21,E}\right] + \lambda_{22,F}\left[g_2\left(q_{2,A}\right) + g_6\left(q_{6,B}\right) + g_{10}\left(q_{10,C}\right) + g_{14}\left(q_{14,D}\right) - q_{22,F}\right] \\
&+\lambda_{23,G}\left[g_3\left(q_{3,A}\right) + g_7\left(q_{7,B}\right) + g_{11}\left(q_{11,C}\right) + g_{15}\left(q_{15,D}\right) - q_{23,G}\right] + \lambda_{24,H}\left[g_4\left(q_{4,A}\right) + g_8\left(q_{8,B}\right) + g_{12}\left(q_{12,C}\right) + g_{16}\left(q_{16,D}\right) - q_{24,H}\right] \\
&-\mu_{1,A} q_{1,A} - \mu_{2,A} q_{2,A} - \mu_{3,A} q_{3,A} - \mu_{4,A} q_{4,A} - \mu_{5,B} q_{5,B} - \mu_{6,B} q_{6,B} - \mu_{7,B} q_{7,B} - \mu_{8,B} q_{8,B} \\
&-\mu_{9,C} q_{9,C} - \mu_{10,C} q_{10,C} - \mu_{11,C} q_{11,C} - \mu_{12,C} q_{12,C} - \mu_{13,D} q_{13,D} - \mu_{14,D} q_{14,D} - \mu_{15,D} q_{15,D} - \mu_{16,D} q_{16,D}
\end{aligned}
$$

The portion of the Lagrangian below the dotted line is, interestingly, the Lagrangian for the following transportation cost minimization problem:

$$
\begin{aligned}
&TC\left(q_{A,17}, q_{B,18}, q_{C,19}, q_{D,20}, q_{21,E}, q_{22,F}, q_{23,G}, q_{24,H}\right) \\
&\triangleq \\
&\text{MIN} \quad VOC_1 q_{1,A} + VOC_2 q_{2,A} + VOC_3 q_{3,A} + VOC_4 q_{4,A} + VOC_5 q_{5,B} + VOC_6 q_{6,B} + VOC_7 q_{7,B} + VOC_8 q_{8,B} \\
&\qquad + VOC_9 q_{9,C} + VOC_{10} q_{10,C} + VOC_{11} q_{11,C} + VOC_{12} q_{12,C} + VOC_{13} q_{13,D} + VOC_{14} q_{14,D} + VOC_{15} q_{15,D} + VOC_{16} q_{16,D}
\end{aligned}
$$

$$
\begin{aligned}
&\text{SUBJECT TO} \\
&q_{A,17} - q_{1,A} - q_{2,A} - q_{3,A} - q_{4,A} = 0 \\
&q_{B,18} - q_{5,B} - q_{6,B} - q_{7,B} - q_{8,B} = 0 \\
&q_{C,19} - q_{9,C} - q_{10,C} - q_{11,C} - q_{12,C} = 0 \\
&q_{D,20} - q_{13,D} - q_{14,D} - q_{15,D} - q_{16,D} = 0 \\
&g_1\left(q_{1,A}\right) + g_5\left(q_{5,B}\right) + g_9\left(q_{9,C}\right) + g_{13}\left(q_{13,D}\right) - q_{21,E} = 0 \\
&g_2\left(q_{2,A}\right) + g_6\left(q_{6,B}\right) + g_{10}\left(q_{10,C}\right) + g_{14}\left(q_{14,D}\right) - q_{22,F} = 0 \\
&g_3\left(q_{3,A}\right) + g_7\left(q_{7,B}\right) + g_{11}\left(q_{11,C}\right) + g_{15}\left(q_{15,D}\right) - q_{23,G} = 0 \\
&g_4\left(q_{4,A}\right) + g_8\left(q_{8,B}\right) + g_{12}\left(q_{12,C}\right) + g_{16}\left(q_{16,D}\right) - q_{24,H} = 0 \\
&q_{1,A} \ge 0;\ q_{2,A} \ge 0;\ q_{3,A} \ge 0;\ q_{4,A} \ge 0;\ q_{5,B} \ge 0;\ q_{6,B} \ge 0;\ q_{7,B} \ge 0;\ q_{8,B} \ge 0; \\
&q_{9,C} \ge 0;\ q_{10,C} \ge 0;\ q_{11,C} \ge 0;\ q_{12,C} \ge 0;\ q_{13,D} \ge 0;\ q_{14,D} \ge 0;\ q_{15,D} \ge 0;\ q_{16,D} \ge 0
\end{aligned}
$$

We know, therefore, that the original monolithic global welfare maximization problem can be written in two parts. The first part of the problem is the foregoing cost minimization problem just posed in green, and the overall monolithic global welfare maximization problem is the following:

$$
\boxed{
\begin{aligned}
&\text{MAX} \quad CS_{17}\left(q_{A,17}\right) + CS_{18}\left(q_{B,18}\right) + CS_{19}\left(q_{C,19}\right) + CS_{20}\left(q_{D,20}\right) \\
&\qquad - TC_{21}\left(q_{21,E}\right) - TC_{22}\left(q_{22,F}\right) - TC_{23}\left(q_{23,G}\right) - TC_{24}\left(q_{24,H}\right) \\
&\qquad - TC\left(q_{A,17}, q_{B,18}, q_{C,19}, q_{D,20}, q_{21,E}, q_{22,F}, q_{23,G}, q_{24,H}\right) \\
&\text{SUBJECT TO} \\
&q_{A,17} \ge 0;\ q_{B,18} \ge 0;\ q_{C,19} \ge 0;\ q_{D,20} \ge 0;\ q_{21,E} \ge 0;\ q_{22,F} \ge 0;\ q_{23,G} \ge 0;\ q_{24,H} \ge 0;
\end{aligned}
}
$$

This presages that the transportation cost minimization problem is not sufficient to solve the whole problem, and yet the **linear programming models ubiquitous in the market are nothing but transportation cost minimization models**. Commercial linear programming models (e.g., GPCM) are cost minimization problems, implicitly derived from monolithic global welfare maximization and "decomposed" out of the more general problem in the way we have just made explicit. This is irrefutable; the mathematics is clear. In the discussion to follow, we will use the green color for the cost minimization problem equations, and we will use the pink color for the monolithic global welfare maximization problem equations.

Based on this two part decomposition, let us first solve the cost minimization problem and then substitute that solution into the overall monolithic global welfare maximization problem.

4.3.1 Kuhn-Tucker Conditions for the Embedded Cost Minimization Problem

The Lagrangian of the embedded cost minimization problem (whose incarnation in the linear input-output case is linear programming) is the following:

$$
\begin{aligned}
L = {} & VOC_1 q_{1,A} + VOC_2 q_{2,A} + VOC_3 q_{3,A} + VOC_4 q_{4,A} + VOC_5 q_{5,B} + VOC_6 q_{6,B} + VOC_7 q_{7,B} + VOC_8 q_{8,B} \\
& + VOC_9 q_{9,C} + VOC_{10} q_{10,C} + VOC_{11} q_{11,C} + VOC_{12} q_{12,C} + VOC_{13} q_{13,D} + VOC_{14} q_{14,D} + VOC_{15} q_{15,D} + VOC_{16} q_{16,D} \\
& + \lambda_{A,17}\left[q_{A,17} - q_{1,A} - q_{2,A} - q_{3,A} - q_{4,A}\right] + \lambda_{B,18}\left[q_{B,18} - q_{5,B} - q_{6,B} - q_{7,B} - q_{8,B}\right] \\
& + \lambda_{C,19}\left[q_{C,19} - q_{9,C} - q_{10,C} - q_{11,C} - q_{12,C}\right] + \lambda_{D,20}\left[q_{D,20} - q_{13,D} - q_{14,D} - q_{15,D} - q_{16,D}\right] \\
& + \lambda_{21,E}\left[g_1\left(q_{1,A}\right) + g_5\left(q_{5,B}\right) + g_9\left(q_{9,C}\right) + g_{13}\left(q_{13,D}\right) - q_{21,E}\right] + \lambda_{22,F}\left[g_2\left(q_{2,A}\right) + g_6\left(q_{6,B}\right) + g_{10}\left(q_{10,C}\right) + g_{14}\left(q_{14,D}\right) - q_{22,F}\right] \\
& + \lambda_{23,G}\left[g_3\left(q_{3,A}\right) + g_7\left(q_{7,B}\right) + g_{11}\left(q_{11,C}\right) + g_{15}\left(q_{15,D}\right) - q_{23,G}\right] + \lambda_{24,H}\left[g_4\left(q_{4,A}\right) + g_8\left(q_{8,B}\right) + g_{12}\left(q_{12,C}\right) + g_{16}\left(q_{16,D}\right) - q_{24,H}\right] \\
& - \mu_{1,A} q_{1,A} - \mu_{2,A} q_{2,A} - \mu_{3,A} q_{3,A} - \mu_{4,A} q_{4,A} - \mu_{5,B} q_{5,B} - \mu_{6,B} q_{6,B} - \mu_{7,B} q_{7,B} - \mu_{8,B} q_{8,B} \\
& - \mu_{9,C} q_{9,C} - \mu_{10,C} q_{10,C} - \mu_{11,C} q_{11,C} - \mu_{12,C} q_{12,C} - \mu_{13,D} q_{13,D} - \mu_{14,D} q_{14,D} - \mu_{15,D} q_{15,D} - \mu_{16,D} q_{16,D}
\end{aligned}
$$

Its Kuhn-Tucker conditions are calculated in three parts.

4.3.1.1 Derivatives of Outputs from Transportation Activities (Sixteen)

$$
\begin{aligned}
\frac{\partial L}{\partial q_{1,A}} &= VOC_1 + \lambda_{21,E} g_1'\left(q_{1,A}\right) - \lambda_{A,17} - \mu_{1,A} = 0 \\
\frac{\partial L}{\partial q_{2,A}} &= VOC_2 + \lambda_{22,F} g_2'\left(q_{2,A}\right) - \lambda_{A,17} - \mu_{2,A} = 0 \\
\frac{\partial L}{\partial q_{3,A}} &= VOC_3 + \lambda_{23,G} g_3'\left(q_{3,A}\right) - \lambda_{A,17} - \mu_{3,A} = 0 \\
\frac{\partial L}{\partial q_{4,A}} &= VOC_4 + \lambda_{24,H} g_4'\left(q_{4,A}\right) - \lambda_{A,17} - \mu_{4,A} = 0 \\
\frac{\partial L}{\partial q_{5,B}} &= VOC_5 + \lambda_{21,E} g_5'\left(q_{5,B}\right) - \lambda_{B,18} - \mu_{5,B} = 0
\end{aligned}
$$

$$\frac{\partial L}{\partial q_{6,B}} = VOC_6 + \lambda_{22,F} g_6 '\left(q_{6,B}\right) - \lambda_{B,18} - \mu_{6,B} = 0$$

$$\frac{\partial L}{\partial q_{7,B}} = VOC_7 + \lambda_{23,G} g_7 '\left(q_{7B}\right) - \lambda_{B,18} - \mu_{7,B} = 0$$

$$\frac{\partial L}{\partial q_{8,B}} = VOC_8 + \lambda_{24,H} g_8 '\left(q_{8,B}\right) - \lambda_{B,18} - \mu_{8,B} = 0$$

$$\frac{\partial L}{\partial q_{9,C}} = VOC_9 + \lambda_{21,E} g_9 '\left(q_{9,C}\right) - \lambda_{C,19} - \mu_{9,C} = 0$$

$$\frac{\partial L}{\partial q_{10,C}} = VOC_{10} + \lambda_{22,F} g_{10} '\left(q_{10,C}\right) - \lambda_{C,19} - \mu_{10,C} = 0$$

$$\frac{\partial L}{\partial q_{11,C}} = VOC_{11} + \lambda_{23,G} g_{11} '\left(q_{11,C}\right) - \lambda_{C,19} - \mu_{11,C} = 0$$

$$\frac{\partial L}{\partial q_{12,C}} = VOC_{12} + \lambda_{24,H} g_{12} '\left(q_{12,C}\right) - \lambda_{C,19} - \mu_{12,C} = 0$$

$$\frac{\partial L}{\partial q_{13,D}} = VOC_{13} + \lambda_{21,E} g_{13} '\left(q_{13,D}\right) - \lambda_{D,20} - \mu_{13,D} = 0$$

$$\frac{\partial L}{\partial q_{14,D}} = VOC_{14} + \lambda_{22,F} g_{14} '\left(q_{14,D}\right) - \lambda_{D,20} - \mu_{14,D} = 0$$

$$\frac{\partial L}{\partial q_{15,D}} = VOC_{15} + \lambda_{23,G} g_{15} '\left(q_{15,D}\right) - \lambda_{D,20} - \mu_{15,D} = 0$$

$$\frac{\partial L}{\partial q_{16,D}} = VOC_{16} + \lambda_{24,H} g_{16} '\left(q_{16,D}\right) - \lambda_{D,20} - \mu_{16,D} = 0$$

4.3.1.2 Hub Balance Equations (Eight)

$$\frac{\partial L}{\partial \lambda_{A,17}} = q_{A,17} - q_{1,A} - q_{2,A} - q_{3,A} - q_{4,A} = 0$$

$$\frac{\partial L}{\partial \lambda_{B,18}} = q_{B,18} - q_{5,B} - q_{6,B} - q_{7,B} - q_{8,B} = 0$$

$$\frac{\partial L}{\partial \lambda_{C,19}} = q_{C,19} - q_{9,C} - q_{10,C} - q_{11,C} - q_{12,C} = 0$$

$$\frac{\partial L}{\partial \lambda_{D,20}} = q_{D,20} - q_{13,D} - q_{14,D} - q_{15,D} - q_{16,D} = 0$$

$$\frac{\partial L}{\partial \lambda_{21,E}} = g_1\left(q_{1,A}\right) + g_5\left(q_{5,B}\right) + g_9\left(q_{9,C}\right) + g_{13}\left(q_{13,D}\right) - q_{21,E} = 0$$

$$\frac{\partial L}{\partial \lambda_{22,F}} = g_2\left(q_{2,A}\right) + g_6\left(q_{6,B}\right) + g_{10}\left(q_{10,C}\right) + g_{14}\left(q_{14,D}\right) - q_{22,F} = 0$$

$$\frac{\partial L}{\partial \lambda_{23,G}} = g_3\left(q_{3,A}\right) + g_7\left(q_{7,B}\right) + g_{11}\left(q_{11,C}\right) + g_{15}\left(q_{15,D}\right) - q_{23,G} = 0$$

$$\frac{\partial L}{\partial \lambda_{24,H}} = g_4\left(q_{4,A}\right) + g_8\left(q_{8,B}\right) + g_{12}\left(q_{12,C}\right) + g_{16}\left(q_{16,D}\right) - q_{24,H} = 0$$

4.3.1.3 *Inequality Kuhn-Tucker Conditions*

$$\begin{array}{lll}
\mu_{1,A}q_{1,A} = 0 & \mu_{1,A} \geq 0 & q_{1,A} \geq 0 \\
\mu_{2,A}q_{2,A} = 0 & \mu_{2,A} \geq 0 & q_{2,A} \geq 0 \\
\mu_{3,A}q_{3,A} = 0 & \mu_{3,A} \geq 0 & q_{3,A} \geq 0 \\
\mu_{4,A}q_{4,A} = 0 & \mu_{4,A} \geq 0 & q_{4,A} \geq 0 \\
\mu_{5,B}q_{5,B} = 0 & \mu_{5,B} \geq 0 & q_{5,B} \geq 0 \\
\mu_{6,B}q_{6,B} = 0 & \mu_{6,B} \geq 0 & q_{6,B} \geq 0 \\
\mu_{7,B}q_{7,B} = 0 & \mu_{7,B} \geq 0 & q_{7,B} \geq 0 \\
\mu_{8,B}q_{8,B} = 0 & \mu_{8,B} \geq 0 & q_{8,B} \geq 0 \\
\mu_{9,C}q_{9,C} = 0 & \mu_{9,C} \geq 0 & q_{9,C} \geq 0 \\
\mu_{10,C}q_{10,C} = 0 & \mu_{10,C} \geq 0 & q_{10,C} \geq 0 \\
\mu_{11,C}q_{11,C} = 0 & \mu_{11,C} \geq 0 & q_{11,C} \geq 0 \\
\mu_{12,C}q_{12,C} = 0 & \mu_{12,C} \geq 0 & q_{12,C} \geq 0 \\
\mu_{13,D}q_{13,D} = 0 & \mu_{13,D} \geq 0 & q_{13,D} \geq 0 \\
\mu_{14,D}q_{14,D} = 0 & \mu_{14,D} \geq 0 & q_{14,D} \geq 0 \\
\mu_{15,D}q_{15,D} = 0 & \mu_{15,D} \geq 0 & q_{15,D} \geq 0 \\
\mu_{16,D}q_{16,D} = 0 & \mu_{16,D} \geq 0 & q_{16,D} \geq 0
\end{array}$$

These are the Kuhn-Tucker conditions for the embedded cost minimization problem, and they completely specify the minimum transportation cost to move fixed supplies to fixed demands.

We can write these equations in complementarity form by combining the derivatives and the inequality constraints:

$$\begin{array}{lll}
\left[VOC_1 + \lambda_{21,E}g_1'\left(q_{1,A}\right) - \lambda_{A,17}\right]q_{1,A} = 0 & VOC_1 + \lambda_{21,E}g_1'\left(q_{1,A}\right) - \lambda_{A,17} \geq 0 & q_{1,A} \geq 0 \\
\left[VOC_2 + \lambda_{22,F}g_2'\left(q_{2,A}\right) - \lambda_{A,17}\right]q_{2,A} = 0 & VOC_2 + \lambda_{22,F}g_2'\left(q_{2,A}\right) - \lambda_{A,17} \geq 0 & q_{2,A} \geq 0 \\
\left[VOC_3 + \lambda_{23,G}g_3'\left(q_{3,A}\right) - \lambda_{A,17}\right]q_{3,A} = 0 & VOC_3 + \lambda_{23,G}g_3'\left(q_{3,A}\right) - \lambda_{A,17} \geq 0 & q_{3,A} \geq 0 \\
\left[VOC_4 + \lambda_{24,H}g_4'\left(q_{4,A}\right) - \lambda_{A,17}\right]q_{4,A} = 0 & VOC_4 + \lambda_{24,H}g_4'\left(q_{4,A}\right) - \lambda_{A,17} \geq 0 & q_{4,A} \geq 0 \\
\left[VOC_5 + \lambda_{21,E}g_5'\left(q_{5,B}\right) - \lambda_{B,18}\right]q_{5,B} = 0 & VOC_5 + \lambda_{21,E}g_5'\left(q_{5,B}\right) - \lambda_{B,18} \geq 0 & q_{5,B} \geq 0 \\
\left[VOC_6 + \lambda_{22,F}g_6'\left(q_{6,B}\right) - \lambda_{B,18}\right]q_{6,B} = 0 & VOC_6 + \lambda_{22,F}g_6'\left(q_{6,B}\right) - \lambda_{B,18} \geq 0 & q_{6,B} \geq 0 \\
\left[VOC_7 + \lambda_{23,G}g_7'\left(q_{7B}\right) - \lambda_{B,18}\right]q_{7B} = 0 & VOC_7 + \lambda_{23,G}g_7'\left(q_{7B}\right) - \lambda_{B,18} \geq 0 & q_{7B} \geq 0 \\
\left[VOC_8 + \lambda_{24,H}g_8'\left(q_{8,B}\right) - \lambda_{B,18}\right]q_{8,B} = 0 & VOC_8 + \lambda_{24,H}g_8'\left(q_{8,B}\right) - \lambda_{B,18} \geq 0 & q_{8,B} \geq 0 \\
\left[VOC_9 + \lambda_{21,E}g_9'\left(q_{9,C}\right) - \lambda_{C,19}\right]q_{9,C} = 0 & VOC_9 + \lambda_{21,E}g_9'\left(q_{9,C}\right) - \lambda_{C,19} \geq 0 & q_{9,C} \geq 0 \\
\left[VOC_{10} + \lambda_{22,F}g_{10}'\left(q_{10,C}\right) - \lambda_{C,19}\right]q_{10,C} = 0 & VOC_{10} + \lambda_{22,F}g_{10}'\left(q_{10,C}\right) - \lambda_{C,19} \geq 0 & q_{10,C} \geq 0 \\
\left[VOC_{11} + \lambda_{23,G}g_{11}'\left(q_{11,C}\right) - \lambda_{C,19}\right]q_{11,C} = 0 & VOC_{11} + \lambda_{23,G}g_{11}'\left(q_{11,C}\right) - \lambda_{C,19} \geq 0 & q_{11,C} \geq 0
\end{array}$$

$$\left[VOC_{12}+\lambda_{24,H}g_{12}'\left(q_{12,C}\right)-\lambda_{C,19}\right]q_{12,C}=0 \quad VOC_{12}+\lambda_{24,H}g_{12}'\left(q_{12,C}\right)-\lambda_{C,19}\geq 0 \quad q_{12,C}\geq 0$$
$$\left[VOC_{13}+\lambda_{21,E}g_{13}'\left(q_{13,D}\right)-\lambda_{D,20}\right]q_{13,D}=0 \quad VOC_{13}+\lambda_{21,E}g_{13}'\left(q_{13,D}\right)-\lambda_{D,20}\geq 0 \quad q_{13,D}\geq 0$$
$$\left[VOC_{14}+\lambda_{22,F}g_{14}'\left(q_{14,D}\right)-\lambda_{D,20}\right]q_{14,D}=0 \quad VOC_{14}+\lambda_{22,F}g_{14}'\left(q_{14,D}\right)-\lambda_{D,20}\geq 0 \quad q_{14,D}\geq 0$$
$$\left[VOC_{15}+\lambda_{23,G}g_{15}'\left(q_{15,D}\right)-\lambda_{D,20}\right]q_{15,D}=0 \quad VOC_{15}+\lambda_{23,G}g_{15}'\left(q_{15,D}\right)-\lambda_{D,20}\geq 0 \quad q_{15,D}\geq 0$$
$$\left[VOC_{16}+\lambda_{24,H}g_{16}'\left(q_{16,D}\right)-\lambda_{D,20}\right]q_{16,D}=0 \quad VOC_{16}+\lambda_{24,H}g_{16}'\left(q_{16,D}\right)-\lambda_{D,20}\geq 0 \quad q_{16,D}\geq 0$$
$$q_{A,17}-q_{1,A}-q_{2,A}-q_{3,A}-q_{4,A}=0$$
$$q_{B,18}-q_{5,B}-q_{6,B}-q_{7,B}-q_{8,B}=0$$
$$q_{C,19}-q_{9,C}-q_{10,C}-q_{11,C}-q_{12,C}=0$$
$$q_{D,20}-q_{13,D}-q_{14,D}-q_{15,D}-q_{16,D}=0$$
$$g_1\left(q_{1,A}\right)+g_5\left(q_{5,B}\right)+g_9\left(q_{9,C}\right)+g_{13}\left(q_{13,D}\right)-q_{21,E}=0$$
$$g_2\left(q_{2,A}\right)+g_6\left(q_{6,B}\right)+g_{10}\left(q_{10,C}\right)+g_{14}\left(q_{14,D}\right)-q_{22,F}=0$$
$$g_3\left(q_{3,A}\right)+g_7\left(q_{7,B}\right)+g_{11}\left(q_{11,C}\right)+g_{15}\left(q_{15,D}\right)-q_{23,G}=0$$
$$g_4\left(q_{4,A}\right)+g_8\left(q_{8,B}\right)+g_{12}\left(q_{12,C}\right)+g_{16}\left(q_{16,D}\right)-q_{24,H}=0$$

This is consistent with the format Gabriel op. cit. assert will emanate from an optimization problem. These 24 conditions will solve for the 24 unknowns—the 16 quantities flowing out of the transportation processes and the 8 Lagrange multipliers, which will be interpreted as prices. As we shall see, this is far too hard and cumbersome to implement and solve. There is a much more efficient way than complementarity to get the same solution, as we will see in Section 5, and it comes from substitution rather than numerical calculation, as we will see in Section 7.

4.3.2 Kuhn-Tucker Conditions for the Monolithic Global Welfare Maximization Problem

Let's write the Kuhn-Tucker conditions for the overall monolithic global welfare maximization problem. First, we pose it in Luenberger form:

$$\begin{aligned}
\text{MIN}\quad & -CS_{17}\left(q_{A,17}\right)-CS_{18}\left(q_{B,18}\right)-CS_{19}\left(q_{C,19}\right)-CS_{20}\left(q_{D,20}\right)\\
& +TC_{21}\left(q_{21,E}\right)+TC_{22}\left(q_{22,F}\right)+TC_{23}\left(q_{23,G}\right)+TC_{24}\left(q_{24,H}\right)\\
& +TC\left(q_{A,17},q_{B,18},q_{C,19},q_{D,20},q_{21,E},q_{22,F},q_{23,G},q_{24,H}\right)\\
\text{SUBJECT TO}\quad & \\
& q_{A,17}\geq 0;\ q_{B,18}\geq 0;\ q_{C,19}\geq 0;\ q_{D,20}\geq 0;\ q_{21,E}\geq 0;\ q_{22,F}\geq 0;\ q_{23,G}\geq 0;\ q_{24,H}\geq 0;
\end{aligned}$$

Then write the Lagrangian, using the notation ω to denote the Lagrange multiplier:

$$
\begin{aligned}
&L = -CS_{17}\left(q_{A,17}\right) - CS_{18}\left(q_{B,18}\right) - CS_{19}\left(q_{C,19}\right) - CS_{20}\left(q_{D,20}\right) \\
&+TC_{21}\left(q_{21,E}\right) + TC_{22}\left(q_{22,F}\right) + TC_{23}\left(q_{23,G}\right) + TC_{24}\left(q_{24,H}\right) \\
&+TC\left(q_{A,17}, q_{B,18}, q_{C,19}, q_{D,20}, q_{21,E}, q_{22,F}, q_{23,G}\right) \\
&-\omega_{A,17}q_{A,17} - \omega_{B,18}q_{B,18} - \omega_{C,19}q_{C,19} - \omega_{D,20}q_{D,20} - \omega_{21,E}q_{21,E} - \omega_{22,F}q_{22,F} - \omega_{23,G}q_{23,G} - \omega_{24,H}q_{24,H}
\end{aligned}
$$

We then take the appropriate partial derivatives and set them to 0 to compute the Kuhn-Tucker conditions:

$$
\begin{aligned}
\frac{\partial L}{\partial q_{A,17}} &= -p_{17}\left(q_{A,17}\right) + \frac{\partial TC\left(q_{A,17}, q_{B,18}, q_{C,19}, q_{D,20}, q_{21,E}, q_{22,F}, q_{23,G}, q_{24,H}\right)}{\partial q_{A,17}} - \omega_{A,17} = 0 \\
\frac{\partial L}{\partial q_{B,18}} &= -p_{18}\left(q_{B,18}\right) + \frac{\partial TC\left(q_{A,17}, q_{B,18}, q_{C,19}, q_{D,20}, q_{21,E}, q_{22,F}, q_{23,G}, q_{24,H}\right)}{\partial q_{B,18}} - \omega_{B,18} = 0 \\
\frac{\partial L}{\partial q_{C,19}} &= -p_{19}\left(q_{C,19}\right) + \frac{\partial TC\left(q_{A,17}, q_{B,18}, q_{C,19}, q_{D,20}, q_{21,E}, q_{22,F}, q_{23,G}, q_{24,H}\right)}{\partial q_{C,19}} - \omega_{C,19} = 0 \\
\frac{\partial L}{\partial q_{D,20}} &= -p_{20}\left(q_{D,20}\right) + \frac{\partial TC\left(q_{A,17}, q_{B,18}, q_{C,19}, q_{D,20}, q_{21,E}, q_{22,F}, q_{23,G}, q_{24,H}\right)}{\partial q_{D,20}} - \omega_{D,20} = 0 \\
\frac{\partial L}{\partial q_{21,E}} &= MC_{21}\left(q_{21,E}\right) + \frac{\partial TC\left(q_{A,17}, q_{B,18}, q_{C,19}, q_{D,20}, q_{21,E}, q_{22,F}, q_{23,G}, q_{24,H}\right)}{\partial q_{21,E}} - \omega_{21,E} = 0 \\
\frac{\partial L}{\partial q_{22,F}} &= MC_{22}\left(q_{22,F}\right) + \frac{\partial TC\left(q_{A,17}, q_{B,18}, q_{C,19}, q_{D,20}, q_{21,E}, q_{22,F}, q_{23,G}, q_{24,H}\right)}{\partial q_{22,F}} - \omega_{22,F} = 0 \\
\frac{\partial L}{\partial q_{23,G}} &= MC_{23}\left(q_{23,G}\right) + \frac{\partial TC\left(q_{A,17}, q_{B,18}, q_{C,19}, q_{D,20}, q_{21,E}, q_{22,F}, q_{23,G}, q_{24,H}\right)}{\partial q_{23,G}} - \omega_{23,G} = 0 \\
\frac{\partial L}{\partial q_{24,H}} &= MC_{24}\left(q_{24,H}\right) + \frac{\partial TC\left(q_{A,17}, q_{B,18}, q_{C,19}, q_{D,20}, q_{21,E}, q_{22,F}, q_{23,G}, q_{24,H}\right)}{\partial q_{24,H}} - \omega_{24,H} = 0
\end{aligned}
$$

4.3.2.1 Summary Proof of the Envelope Theorem to Solve for the Derivatives of Total Cost

The Envelope Theorem can be used to find the derivatives of total cost. We will give a quick proof in one dimension and then apply the envelope theorem. (This proof in one dimension is easily extendable to multiple dimensions.) The Envelope Theorem begins with the constrained optimization problem:

$$
\begin{aligned}
&\text{MIN} \quad f\left(x\right) \\
&\text{SUBJECT TO} \quad h\left(c, x\right) = 0
\end{aligned}
$$

The solution x* is obviously a function of the parameter c. We denote the solution x*(c). The Lagrangian for this problem is:

$$L\left(x, \lambda\right) = f\left(x\right) + \lambda h\left(c, x\right)$$

The Kuhn-Tucker conditions for this problem, which must hold for the solution x*(c), are:

$$\frac{\partial L(x,\lambda)}{\partial x} = \frac{\partial f(x)}{\partial x} + \lambda \frac{\partial h(c,x)}{\partial x} = 0$$
$$\frac{\partial L(x,\lambda)}{\partial x} = h(c,x) = 0$$

Equation 1

We want to know what the sensitivity of the optimal value is with respect to the parameter c.

$$\frac{\partial f\left[x^*(c)\right]}{\partial c}$$

Equation 2

We expand this sensitivity expression using the chain rule:

$$\frac{\partial f\left[x^*(c)\right]}{\partial c} = \frac{\partial f\left[x^*(c)\right]}{\partial x}\frac{\partial x^*(c)}{\partial c}$$

The first Kuhn-Tucker condition in Equation 1 is:

$$\frac{\partial L(x,\lambda)}{\partial x} = \frac{\partial f(x)}{\partial x} + \lambda \frac{\partial h(c,x)}{\partial x} = 0 \Rightarrow \frac{\partial f(x)}{\partial x} = -\lambda \frac{\partial h(c,x)}{\partial x}$$

We substitute the optimal answer x*(c) to write:

$$\frac{\partial f\left[x^*(c)\right]}{\partial x} = -\lambda \frac{\partial h\left[c, x^*(c)\right]}{\partial x}$$

We substitute into the expression for the sensitivity in Equation 2:

$$\frac{\partial f\left[x^*(c)\right]}{\partial c} = -\lambda \frac{\partial h\left[c, x^*(c)\right]}{\partial x}\frac{\partial x^*(c)}{\partial c}$$

Equation 3

We then write the constraint at the optimum answer:

$$h\left[c, x^*(c)\right] = 0$$

We differentiate with respect to c:

$$\frac{\partial h\left[c, x^*(c)\right]}{\partial c} + \frac{\partial h\left[c, x^*(c)\right]}{\partial x}\frac{\partial x^*(c)}{\partial c} = 0$$

We rearrange to isolate the derivative of the optimum with respect to c:

$$\frac{\partial x^*(c)}{\partial c} = -\frac{\dfrac{\partial h\left[c, x^*(c)\right]}{\partial c}}{\dfrac{\partial h\left[c, x^*(c)\right]}{\partial x}}$$

We substitute this into the desired sensitivity in Equation 3 above to write:

$$\frac{\partial f\left[x^*(c)\right]}{\partial c} = -\lambda \frac{\partial h\left[c, x^*(c)\right]}{\partial x} \frac{\partial x^*(c)}{\partial c} = \lambda \frac{\partial h\left[c, x^*(c)\right]}{\partial c}$$

By this theorem, the derivative of an optimized objective function with respect to the constant c is the derivative of the constraint containing this constant (ensuring the correct sign and magnitude) times the Lagrange multiplier for the constraint. This is a powerful result, allowing us to calculate the sensitivity of a solution to a parameter.

4.3.2.2 Solve the Equations Using the Envelope Theorem

By the envelope theorem just proven:

$$\frac{\partial TC\left(q_{A,17}, q_{B,18}, q_{C,19}, q_{D,20}, q_{21,E}, q_{22,F}, q_{23,G}, q_{24,H}\right)}{\partial q_{A,17}} = \lambda_{A,17}$$
$$\frac{\partial TC\left(q_{A,17}, q_{B,18}, q_{C,19}, q_{D,20}, q_{21,E}, q_{22,F}, q_{23,G}, q_{24,H}\right)}{\partial q_{B,18}} = \lambda_{B,18}$$
$$\frac{\partial TC\left(q_{A,17}, q_{B,18}, q_{C,19}, q_{D,20}, q_{21,E}, q_{22,F}, q_{23,G}, q_{24,H}\right)}{\partial q_{C,19}} = \lambda_{C,19}$$
$$\frac{\partial TC\left(q_{A,17}, q_{B,18}, q_{C,19}, q_{D,20}, q_{21,E}, q_{22,F}, q_{23,G}, q_{24,H}\right)}{\partial q_{D,20}} = \lambda_{D,20}$$
$$\frac{\partial TC\left(q_{A,17}, q_{B,18}, q_{C,19}, q_{D,20}, q_{21,E}, q_{22,F}, q_{23,G}, q_{24,H}\right)}{\partial q_{21,E}} = -\lambda_{21,E}$$
$$\frac{\partial TC\left(q_{A,17}, q_{B,18}, q_{C,19}, q_{D,20}, q_{21,E}, q_{22,F}, q_{23,G}, q_{24,H}\right)}{\partial q_{22,F}} = -\lambda_{22,F}$$
$$\frac{\partial TC\left(q_{A,17}, q_{B,18}, q_{C,19}, q_{D,20}, q_{21,E}, q_{22,F}, q_{23,G}, q_{24,H}\right)}{\partial q_{23,G}} = -\lambda_{23,G}$$
$$\frac{\partial TC\left(q_{A,17}, q_{B,18}, q_{C,19}, q_{D,20}, q_{21,E}, q_{22,F}, q_{23,G}, q_{24,H}\right)}{\partial q_{24,H}} = -\lambda_{24,H}$$

Substitution gives the desired Kuhn-Tucker conditions, which are precisely consistent with the original monolithic global welfare maximization problem.

$$\frac{\partial L}{\partial q_{A,17}} = -p_{17}\left(q_{A,17}\right) + \lambda_{A,17} - \omega_{A,17} = 0$$
$$\frac{\partial L}{\partial q_{B,18}} = -p_{18}\left(q_{B,18}\right) + \lambda_{B,18} - \omega_{B,18} = 0$$
$$\frac{\partial L}{\partial q_{C,19}} = -p_{19}\left(q_{C,19}\right) + \lambda_{C,19} - \omega_{C,19} = 0$$
$$\frac{\partial L}{\partial q_{D,20}} = -p_{20}\left(q_{D,20}\right) + \lambda_{D,20} - \omega_{D,20} = 0$$
$$\frac{\partial L}{\partial q_{21,E}} = MC_{21}\left(q_{21,E}\right) - \lambda_{21,E} - \omega_{21,E} = 0$$
$$\frac{\partial L}{\partial q_{22,F}} = MC_{22}\left(q_{22,F}\right) - \lambda_{22,F} - \omega_{22,F} = 0$$
$$\frac{\partial L}{\partial q_{23,G}} = MC_{23}\left(q_{23,G}\right) - \lambda_{23,G} - \omega_{23,G} = 0$$
$$\frac{\partial L}{\partial q_{24,H}} = MC_{24}\left(q_{24,H}\right) - \lambda_{24,H} - \omega_{24,H} = 0$$

We must append additional equations for the inequality constraints, which are the following:

$$\begin{array}{lll}
\omega_{A,17} q_{A,17} = 0 & \omega_{A,17} \geq 0 & q_{A,17} \geq 0 \\
\omega_{B,18} q_{B,18} = 0 & \omega_{B,18} \geq 0 & q_{B,18} \geq 0 \\
\omega_{C,19} q_{C,19} = 0 & \omega_{C,19} \geq 0 & q_{C,19} \geq 0 \\
\omega_{D,20} q_{D,20} = 0 & \omega_{D,20} \geq 0 & q_{D,20} \geq 0 \\
\omega_{21,E} q_{21,E} = 0 & \omega_{21,E} \geq 0 & q_{21,E} \geq 0 \\
\omega_{22,F} q_{22,F} = 0 & \omega_{22,F} \geq 0 & q_{22,F} \geq 0 \\
\omega_{23,G} q_{23,G} = 0 & \omega_{23,G} \geq 0 & q_{23,G} \geq 0 \\
\omega_{24,H} q_{24,H} = 0 & \omega_{24,H} \geq 0 & q_{24,H} \geq 0
\end{array}$$

We can write this in complementarity form:

$$\begin{array}{lll}
\left[\lambda_{A,17} - p_{17}\left(q_{A,17}\right)\right] q_{A,17} = 0 & \lambda_{A,17} - p_{17}\left(q_{A,17}\right) \geq 0 & q_{A,17} \geq 0 \\
\left[\lambda_{B,18} - p_{18}\left(q_{B,18}\right)\right] q_{B,18} = 0 & \lambda_{B,18} - p_{18}\left(q_{B,18}\right) \geq 0 & q_{B,18} \geq 0 \\
\left[\lambda_{C,19} - p_{19}\left(q_{C,19}\right)\right] q_{C,19} = 0 & \lambda_{C,19} - p_{19}\left(q_{C,19}\right) \geq 0 & q_{C,19} \geq 0 \\
\left[\lambda_{D,20} - p_{20}\left(q_{D,20}\right)\right] q_{D,20} = 0 & \lambda_{D,20} - p_{20}\left(q_{D,20}\right) \geq 0 & q_{D,20} \geq 0 \\
\left[MC_{21}\left(q_{21,E}\right) - \lambda_{21,E}\right] q_{21,E} = 0 & MC_{21}\left(q_{21,E}\right) - \lambda_{21,E} \geq 0 & q_{21,E} \geq 0 \\
\left[MC_{22}\left(q_{22,F}\right) - \lambda_{22,F}\right] q_{22,F} = 0 & MC_{22}\left(q_{22,F}\right) - \lambda_{22,F} \geq 0 & q_{22,F} \geq 0 \\
\left[MC_{23}\left(q_{23,G}\right) - \lambda_{23,G}\right] q_{23,G} = 0 & MC_{23}\left(q_{23,G}\right) - \lambda_{23,G} \geq 0 & q_{23,G} \geq 0 \\
\left[MC_{24}\left(q_{24,H}\right) - \lambda_{24,H}\right] q_{24,H} = 0 & MC_{24}\left(q_{24,H}\right) - \lambda_{24,H} \geq 0 & q_{24,H} \geq 0
\end{array}$$

The linear program is embedded within a complementarity problem of its own! The linear programming problem must solve for the Lagrange multipliers λ, and the complementarity equation thereafter solves for the quantities and its own Lagrange multipliers ω. Then the whole problem must be iterated until both problems give completely unchanging answers. This is a difficult problem to solve, a problem we avoid with network microeconomic equilibrium, as we shall see in Section 5 below. (It is a problem that linear programming people choose to avoid simply by ignoring it. That is not appropriate.)

4.3.3 Decomposition and Isolation of the Transportation Problem from the Supply and Demand Problem

The attractive part of this two part decomposition is that we can isolate the transportation problem away from the supply and demand problem. We can solve the transportation problem assuming fixed demand and fixed supply and thereafter insert the answer into the supply and demand problem and iterate between the two. That would allow EIA to deploy a different solution technique for the transportation and logistics portion of the problem than for the supply and demand portion. This decomposition and breakout of transportation away from supply and demand is very propitious, for it allows hybrid methods, an efficient method to deal with transportation and a same or different, efficient method to deal with supply and demand. It does not doom your architecture to one huge "solver."

This decomposition also puts to rest the myth that a linear programming solution is a full solution to the market problem. It is **not**. All it solves is the embedded transportation "network" problem, which we have characterized as a transportation problem here. Linear programming solves a piece of the problem but not the whole problem. Any assertion that linear programming is a solution to the economic problem has in this section been proven mathematically to be incorrect, claims of linear programming vendors notwithstanding.

4.3.4 From Whence Does the Linear Program Devolve That EIA Has Used in the Past?

The linear programming representation is derived from the embedded cost minimization problem we derived in the previous section, but with all input-output relationships being Leontief. It is the following[24]:

$$TC\left(q_{A,17}, q_{B,18}, q_{C,19}, q_{D,20}, q_{21,E}, q_{22,F}, q_{23,G}, q_{24,H}\right) \triangleq$$

$$\begin{aligned} \text{MIN} \quad & VOC_1 q_{1,A} + VOC_2 q_{2,A} + VOC_3 q_{3,A} + VOC_4 q_{4,A} + VOC_5 q_{5,B} + VOC_6 q_{6,B} + VOC_7 q_{7,B} + VOC_8 q_{8,B} \\ & + VOC_9 q_{9,C} + VOC_{10} q_{10,C} + VOC_{11} q_{11,C} + VOC_{12} q_{12,C} + VOC_{13} q_{13,D} + VOC_{14} q_{14,D} + VOC_{15} q_{15,D} + VOC_{16} q_{16,D} \end{aligned}$$

SUBJECT TO

$$q_{1,A} + q_{2,A} + q_{3,A} + q_{4,A} = q_{A,17}$$

$$q_{5,B} + q_{6,B} + q_{7,B} + q_{8,B} = q_{B,18}$$

[24] The final four equality constraints have been multiplied by -1. Nothing is affected by making such a multiplication. This multiplication by -1 will put the linear programming problem into the standard Koopmans-Hitchcock activity analysis format, which has proven to be so insightful in the past.

$$q_{9,C} + q_{10,C} + q_{11,C} + q_{12,C} = q_{C,19}$$

$$q_{13,D} + q_{14,D} + q_{15,D} + q_{16,D} = q_{D,20}$$

$$-\frac{q_{1,A}}{\eta_1} - \frac{q_{5,B}}{\eta_5} - \frac{q_{9,C}}{\eta_9} - \frac{q_{13,D}}{\eta_{13}} = -q_{21,E}$$

$$-\frac{q_{2,A}}{\eta_2} - \frac{q_{6,B}}{\eta_6} - \frac{q_{10,C}}{\eta_{10}} - \frac{q_{14,D}}{\eta_{14}} = -q_{22,F}$$

$$-\frac{q_{3,A}}{\eta_3} - \frac{q_{7,B}}{\eta_7} - \frac{q_{11,C}}{\eta_{11}} - \frac{q_{15,D}}{\eta_{15}} = -q_{23,G}$$

$$-\frac{q_{4,A}}{\eta_4} - \frac{q_{8,B}}{\eta_8} - \frac{q_{12,C}}{\eta_{12}} - \frac{q_{16,D}}{\eta_{16}} = -q_{24,H}$$

$$q_{1,A} \ge 0 \quad q_{2,A} \ge 0 \quad q_{3,A} \ge 0 \quad q_{4,A} \ge 0$$

$$q_{5,B} \ge 0 \quad q_{6,B} \ge 0 \quad q_{7,B} \ge 0 \quad q_{8,B} \ge 0$$

$$q_{9,C} \ge 0 \quad q_{10,C} \ge 0 \quad q_{11,C} \ge 0 \quad q_{12,C} \ge 0$$

$$q_{13,D} \ge 0 \quad q_{14,D} \ge 0 \quad q_{15,D} \ge 0 \quad q_{16,D} \ge 0$$

The constraint set is written in constraint matrix form, which allows us to see exactly where EIA's linear programming activity analysis would lie (ignoring additional constraints that EIA might deploy for various activities). Figure 8 is the constraint set for the primal linear programming problem.

Figure 8: Constraint Matrix for Primal Linear Programming Problem

		Activities																		
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	$q_{1,A}$		
Hubs	**A**	1	1	1	1													$q_{2,A}$	=	$q_{A,17}$
	B					1	1	1	1									$q_{3,A}$	=	$q_{B,19}$
	C									1	1	1	1					$q_{4,A}$	=	$q_{C,19}$
	D													1	1	1	1	$q_{5,B}$	=	$q_{D,20}$
	E	$-1/\eta_1$				$-1/\eta_5$				$-1/\eta_9$				$-1/\eta_{13}$				$q_{6,B}$	=	$-q_{21,E}$
	F		$-1/\eta_2$				$-1/\eta_6$				$-1/\eta_{10}$				$-1/\eta_{14}$			$q_{7,B}$	=	$-q_{22,F}$
	G			$-1/\eta_3$				$-1/\eta_7$				$-1/\eta_{11}$				$-1/\eta_{15}$		$q_{8,B}$	=	$-q_{23,G}$
	H				$-1/\eta_4$				$-1/\eta_8$				$-1/\eta_{12}$				$-1/\eta_{16}$	$q_{9,C}$	=	$-q_{24,H}$
																		$q_{10,C}$		
																		$q_{11,C}$		
																		$q_{12,C}$		
																		$q_{13,D}$		
																		$q_{14,D}$		
																		$q_{15,D}$		
																		$q_{16,D}$		

The constraint matrix in Figure 8 is written in the Koopmans Hitchcock activity analysis form, an extremely important form from the perspective of interpretation. Each activity (each column) has a negative number at the hub location from which it draws in energy, and it has a positive number in the hub location into which it delivers energy. We can see the complexity of this constraint matrix beginning to rear its head. If EIA has to add a link or delete a link (much less build a world scope transportation matrix), EIA faces terrific matrix management headache. (Assuredly EIA knows how unworkable linear programming has been for large transportation and other problems. No mystery here; this section indicates why from a mathematical perspective. You have to manage a constraint matrix that is much bigger than the problem actually needs, a matrix in which **<u>the sparsity itself IS the problem formulation</u>**. Representing a transportation system is akin to managing the sparsity of a matrix!) You always have to deal with highly sophisticated, highly exacting matrix sparsity. Indeed, this sparsity has intrinsic and inescapable meaning in the Koopmans Hitchcock linear programming context. If it is missed, your formulation is wrong. It cannot be ignored; it has to be dealt with flawlessly. In an intrinsically important sense, this sparsity itself **is** the problem formulation. That is what makes the implementation so difficult.

The full linear program, absent the non-negativity constraints, appears in Figure 9. We can see that the number of rows is equal to the number of hubs in the model, and the number of columns is equal to the number of transportation activities in the model. Koopmans and Hitchcock and Samuelson knew this long ago. This is the "Cartesian product" of hubs times activities idea. This is a very sparse yet large matrix, consistent with Koopmans Hitchcock linear programming activity analysis. The sparsity of the matrix gives us a strong premonition about the brute force nature of the linear programming formulation, particularly for spatial problems like transportation. We will see that the network equilibrium representation will have none of this sparsity and therefore none of this brute force nature. Linear programming faces an 8 by 16 highly sparse matrix just for the transportation portion of the problem in Figure 1. Plus linear programming is intrinsically married to Leontief production functions for transportation. Moreover, thus far, this linear programming representation lacks arbitrary constraints, which are so ubiquitous with linear programming formulations. Plus it does not "linearize" any nonlinear functions, all of which add very substantially to the number of rows and number of columns and thereby to the Cartesian product of rows times columns. This presages the fact that there will be a bare minimum of 16 volumetric flows that are the solution here, accompanied by the 8 dual variables or shadow prices. That is precisely what we have proven previously for the nonlinear complementarity case. However, we shall prove that the complementarity case requires a much bigger Jacobean matrix than this linear programming matrix, which is already substantially too big.

Figure 9: Primal Linear Programming Problem sans Nonnegativity Constraints

	MIN	$VOC_1q_{1,A}$	$VOC_2q_{2,A}$	$VOC_3q_{3,A}$	$VOC_4q_{4,A}$	$VOC_5q_{5,B}$	$VOC_6q_{6,B}$	$VOC_7q_{7,B}$	$VOC_8q_{8,B}$	$VOC_9q_{9,C}$	$VOC_{10}q_{10,C}$	$VOC_{11}q_{11,C}$	$VOC_{12}q_{12,C}$	$VOC_{13}q_{13,D}$	$VOC_{14}q_{14,D}$	$VOC_{15}q_{15,D}$	$VOC_{16}q_{16,D}$			
		Activities																		
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	$q_{1,A}$		
Hubs	A	1	1	1	1													$q_{2,A}$	=	$q_{A,17}$
	B					1	1	1	1									$q_{3,A}$	=	$q_{B,19}$
	C									1	1	1	1					$q_{4,A}$	=	$q_{C,19}$
	D													1	1	1	1	$q_{5,B}$	=	$q_{D,20}$
	E	$-1/\eta_1$				$-1/\eta_5$				$-1/\eta_9$				$-1/\eta_{13}$				$q_{6,B}$	=	$-q_{21,E}$
	F		$-1/\eta_2$				$-1/\eta_6$				$-1/\eta_{10}$				$-1/\eta_{14}$			$q_{7,B}$	=	$-q_{22,F}$
	G			$-1/\eta_3$				$-1/\eta_7$				$-1/\eta_{11}$				$-1/\eta_{15}$		$q_{8,B}$	=	$-q_{23,G}$
	H				$-1/\eta_4$				$-1/\eta_8$				$-1/\eta_{12}$				$-1/\eta_{16}$	$q_{9,C}$	=	$-q_{24,H}$
																		$q_{10,C}$		
																		$q_{11,C}$		
																		$q_{12,C}$		
																		$q_{13,D}$		
																		$q_{14,D}$		
																		$q_{15,D}$		
																		$q_{16,D}$		

EIA's linear programming formulation will have a row for every hub in your model, and it will have a column for every activity in your model. (Additional rows and columns will emerge as you discretize nonlinear curves to incorporate any nonlinearities, a most unsavory practice that has beset linear programming formulations for years.) It is easy to see why it is such a profound headache to construct and manage a monstrous, sparse constraint matrix. EIA certainly knows this very well in its quest to implement a problem the scope of world gas or world oil transportation modeling. EIA can see why a problem that size simply is not viable in light of the magnitude of disaggregation ultimately needed for your world scope transportation model. The problem formulation is large and unwieldy, even in the simple case summarized here. (Think about the complicated **block diagonal** structure that emerges when the time dimension is introduced!)

The dimensionality of this matrix presages precisely how dimensionally intensive the nonlinear complementarity problem truly is, complementarity being the direct cousin of linear programming. Nonlinear complementarity is much larger in terms of dimension than linear programming, and more difficult specifically because of the nonlinearity. In the nonlinear case, the rows would continue to represent hubs and the columns would correspond to the nonlinear input-output relationships. One can but imagine how much worse this all gets when one introduces the temporal dimension (multiple forward time periods). When EIA has thousands of transportation processes and tens of time points, management of this awkwardly structured constraint matrix is literally impossible (not to mention solution time and size). This type of intensity will not serve EIA in its quest for effective, efficient, and representative transportation models at the world scope level.

In spite of all this complexity, we emphasize that this cost minimizing linear programming formulation is by itself **insufficient** to solve the economic problem. It represents only part of the problem. One still must solve the supply side at the upstream edge of the transportation network and solve the demand side at the downstream edge of the transportation network and iterate them with the transportation solution! EIA has to somehow iterate your linear programming solution to this cost minimization problem for transportation with the supply and demand portions of your problem at the boundaries of the transportation system! This embedded, ponderous linear programming formulation is merely the transportation portion of the overall problem. We have just shown mathematically that this linear programming transportation problem is **embedded** within and comprises only a portion of the total problem, a problem that contains the supply curves, the demand curves, and the non-negativity constraints for those supply and demand curves. The transportation problem takes the supplies and the demands as given and minimizes the cost of transportation. Anyone claiming that linear programming representations of transportation are a complete economic solution is wrong. It is only part of the solution.

Luenberger has shown us that the linear programming problem expressed as:

$$\begin{aligned} &\text{MIN} \quad c^T x \\ &\text{Sub. To} \quad Ax = b \\ &\qquad\qquad\; x \geq 0 \end{aligned}$$

corresponds to the dual problem:

$$\begin{aligned} &\text{MAX} \quad b^T \lambda \\ &\text{Sub. To} \quad A^T \lambda \leq c \end{aligned}$$

The dual problem for this example can be derived from the Kuhn-Tucker conditions for the primal problem, but we will dispense with that here. The dual program for the foregoing problem is depicted in full form in Figure 10.

Figure 10: Dual Linear Programming Problem

	MAX	$\lambda_{A,17}q_{A,17}$	$\lambda_{B,18}q_{B,18}$	$\lambda_{C,19}q_{C,19}$	$\lambda_{D,20}q_{D,20}$	$\lambda_{21,E}q_{21,E}$	$\lambda_{22,F}q_{22,F}$	$\lambda_{23,G}q_{23,G}$	$\lambda_{24,H}q_{24,H}$			
		Hubs										
		A	B	C	D	E	F	G	H			
Activities	1	1				$-1/\eta_1$				$\lambda_{A,17}$	$\leq$	VOC_1
	2	1					$-1/\eta_2$			$\lambda_{B,18}$	$\leq$	VOC_2
	3	1						$-1/\eta_3$		$\lambda_{C,19}$	$\leq$	VOC_3
	4	1							$-1/\eta_4$	$\lambda_{D,20}$	$\leq$	VOC_4
	5		1			$-1/\eta_5$				$\lambda_{21,E}$	$\leq$	VOC_5
	6		1				$-1/\eta_6$			$\lambda_{22,F}$	$\leq$	VOC_6
	7		1					$-1/\eta_7$		$\lambda_{23,G}$	$\leq$	VOC_7
	8		1						$-1/\eta_8$	$\lambda_{24,H}$	$\leq$	VOC_8
	9			1		$-1/\eta_9$					$\leq$	VOC_9
	10			1			$-1/\eta_{10}$				$\leq$	VOC_{10}
	11			1				$-1/\eta_{11}$			$\leq$	VOC_{11}
	12			1					$-1/\eta_{12}$		$\leq$	VOC_{12}
	13				1	$-1/\eta_{13}$					$\leq$	VOC_{13}
	14				1		$-1/\eta_{14}$				$\leq$	VOC_{14}
	15				1			$-1/\eta_{15}$			$\leq$	VOC_{15}
	16				1				$-1/\eta_{16}$		$\leq$	VOC_{16}

One actually has to solve both the primal and the dual problem here in order to obtain the transportation flows and the Lagrange multipliers, which are interpreted as prices. (The dual can be derived by writing the Kuhn-Tucker conditions for the original problem and organizing the equations containing the Lagrange multipliers.) The dual variables are the Lagrange multipliers on the constraints of the cost minimization problem.

We designate the solution to the linear programming solution to be:

$$TC\left(q_{A,17}, q_{B,18}, q_{C,19}, q_{D,20}, q_{21,E}, q_{22,F}, q_{23,G}, q_{24,H}\right)$$

Even after we solve the cost minimizing linear programming problem to derive this function, we still must solve the full monolithic global welfare maximization problem that has this linear programming problem embedded within it.

$$\begin{aligned}
&\text{MAX}\\
&CS_{17}\left(q_{A,17}\right)+CS_{18}\left(q_{B,18}\right)+CS_{19}\left(q_{C,19}\right)+CS_{20}\left(q_{D,20}\right)\\
&-TC_{21}\left(q_{21,E}\right)-TC_{22}\left(q_{22,F}\right)-TC_{23}\left(q_{23,G}\right)-TC_{24}\left(q_{24,H}\right)\\
&-TC\left(q_{A,17}, q_{B,18}, q_{C,19}, q_{D,20}, q_{21,E}, q_{22,F}, q_{23,G}, q_{24,H}\right)\\
&\quad \text{SUBJECT TO}\\
&\quad q_{A,17}\geq 0 \quad q_{B,18}\geq 0 \quad q_{C,19}\geq 0 \quad q_{D,20}\geq 0\\
&\quad q_{21,E}\geq 0 \quad q_{22,F}\geq 0 \quad q_{23,G}\geq 0 \quad q_{24,H}\geq 0
\end{aligned}$$

We wrote the Kuhn-Tucker conditions for the nonlinear analog of this problem in the previous section.

EIA has de facto been using a linear programming problem embedded within the foregoing complementarity problem, whether or not that complementarity problem was specifically recognized as such. (Dr. William Hogan in approximately 1979 in a private conversation with Dale Nesbitt noted that the PIES people recognized that their linear programming model was embedded in a larger economic context with nonlinear demand and nonlinear supply. It has long been recognized that the linear program itself is insufficient; it only addresses part of the problem. As we recall, the specific approach to inserting supply and demand was not consistent with this development). The same is true for all the vendors of linear programs; linear programs are not by themselves sufficient to solve the world or regional gas or world or regional oil transportation-supply-demand problems.

When one fully understands the mathematics herein, assuredly no one would formulate a transportation problem using linear programming unless he or she thought there was no other way to solve it. That may well have been the case when NEMS was originated in the 1990s, but it is not the case today.[25] As we shall see in the next section, such is definitely not the case. Quite the contrary, there is a substantially more efficient and direct implementation possible. The architecture and logistics of the linear programming formulation are difficult and irregular, let alone the need to iterate between a complementarity problem and a linear programming problem. One can approach the whole problem directly as we will show below.

The reason we went to all this effort to detail the linear programming is simple. The linear programming algorithm is huge and requires a huge and sparse matrix. It is an expensive, error-prone challenge to manage and formulate. Assuredly the nonlinear analog, the complementary problem, is even larger, and it remains equally sparse. (We do not see how it could possibly be smaller.)

4.4 A Serious Problem—Constraints or Market Imperfections or Other Interventions Distort the Interpretation of Lagrange Multipliers on Constraints as Prices

We can posit a nonlinear monolithic global welfare maximization problem and solve it using a concentrated, highly complex, algorithmically difficult complementarity method. Now, suppose someone decides that one of the transportation processes needs a “constraint,” a “market imperfection,” an arbitrary force applied to that transportation process. Suppose there is some other “modeler intervention.” Suppose in particular, someone put in a constraint on the level of output of one of the transportation nodes. What is the Lagrange multiplier on such a constraint? Is it a price?

The answer is no; it is not a price and not interpretable as a price. It is the derivative of the monolithic global welfare function with respect to one more unit of the constraint. More precisely, it is the sensitivity result we proved in the Envelope Theorem section previously. It is a “shadow price” in the classical sense of the word, and it is not interpretable as a price. Neither are any of the other Lagrange multipliers in the problem, all of which will change by the imposition of the constraint. Indeed, constraints introduce Lagrange multipliers that are:

[25] Network microeconomic equilibrium modeling became commercial in 1997. Complementarity modeling seemed to emerge even later. EIA’s decisions in the early 1990s were probably appropriate for the time, but that would not be appropriate today.

$$\lambda_{\text{Constraint}} = \frac{\partial \text{Monolithic Global Social Welfare}}{\partial \text{Constraint}}$$

Constraints might well have the intended effect on the quantities constrained, but they can destroy the interpretation of Lagrange multipliers as market prices. This is where the notion of monolithic global welfare, integrability of demand, and all those questionable concepts break down. They cannot be overlooked nor swept under the rug. The minute one introduces a constraint or “market imperfection,” or an “art of modeling” equation, the interpretation of Lagrange multipliers as prices can go right out the window. The Lagrange multipliers become the variation in monolithic global welfare with respect to relaxation in the constraint. (If one does not have an integrable demand function, this variation becomes meaningless.) This is highly unsatisfying and in the final analysis completely arbitrary. Every economist knows that there is no such thing as global welfare. There is individual welfare, but there is no such thing as global welfare or group utility or group welfare. We know from the Nobel laureate Arrow and others that there is no “group utility function” or “group welfare function” that respects the individual utility functions of the individual citizens of the world. We know that no organization such as EIA wants its model to be seeking maximization of global welfare because global welfare does not exist and people do not and would not collude to increase global welfare even if it did. To allow even the slightest possibility of global welfare (other than Samuelson’s measure) to enter an economic model solution would be fundamentally dishonest in our view. EIA needs to minimize any possibility of that by eliminating global welfare altogether from the outset.

4.5 Why Not Some Other Global Welfare Function?

Complementarity has carefully selected the assumed-to-be-integrable, Samuelsonian monolithic global welfare function so that its equations represent a competitive market. (We shall see the true implications of this assumption later in this paper.) Let’s be creative here! Why not choose some other world welfare function or world utility function? “Let’s embed CO2 emissions in there!” “Let’s put conservation in there!” “Let’s put land use in there!” “Let’s put fracking bans in there!” “Let’s put preference for energy independence in there!” “Let’s put preference for sanctions against Iran in there!” “Let’s put fuel cell automobiles in there!” It is a short step indeed from the Samuelsonian monolithic global welfare function to an arbitrary monolithic global welfare or utility function that you want all your agents to march to. With monolithic global welfare maximization, it is simply too easy (and too tempting) to impose some arbitrary welfare measure and force the agents to act in collusive lockstep so as to maximize it or to examine that “on a policy basis.” It is tempting to do that to “simulate some policy action” by Federal or state government. That is one of the primary reasons to stay as far away as one can from a monolithic global welfare maximization approach. It is simply too easy to slip into global welfare maximization for its own sake. When you do, the meaning and realism of model solutions are completely out the window, and the honesty of model solutions is compromised.

4.6 Complementarity Solution Algorithms

Complementarity solution algorithms are inescapably iterative; there is no way to guarantee convergence to the monolithic global welfare optimum. That is clear from the operations research literature and from the list of algorithms in Gabriel op. cit., which we will cite below. This section addresses some claims that have been variously made about complementarity algorithms. The section deals with both linear programming (a special case subset of complementarity) and general nonlinear complementarity. We reiterate that we are not impressed by anything less than fully nonlinear complementarity. Partial or wholly linear cases are special cases.

4.6.1 Claims, Myths, and Realities of Complementarity Solution Algorithms

"Yeah, but the linear programming algorithm is reliable; it always gets to an answer." Sure enough, linear programming is a reliable way to get to the wrong answer, a partial, incomplete answer, a subproblem within the main problem. We have already proved previously in Section 4 that linear programming is a partial answer, not the whole answer. That is the first difficulty with the foregoing assertion.

The second difficulty is that, in reality, the linear programming algorithm is not guaranteed to get to an answer; the assertion is untrue on its face. There are situations in which there is "no feasible basis" or in which degeneracy can thwart the algorithm. Linear programming is no "insurance policy" that guarantees an answer.

Third, as we show herein, the linear program must be iterated against upstream supply and downstream demand curves to get to the correct, complete answer, an iteration in which the extremely large linear program must be run iteratively myriad times, embedded as one single equation in a large set of complementarity equations. (As we have shown previously here in Section 4, the complete linear program is but one equation within a large set of complementarity equations to be solved. The whole linear program would have to be one of a large system of complementarity equations.) Isn't that comforting! So what if a linear programming model gets to an answer. The complementarity equations within which it is subsequently embedded have to be solved numerically or iteratively, and there is no guarantee that the larger problem gets to an answer. Even if very large logistical problems could be reliably organized in the form of an almost vanishingly sparse linear programming constraint matrix that is required (which we do not believe to be realistic), then it is true that the simplex algorithm is a pure descent algorithm (in the absence of degeneracy) that moves toward a cost minimizing solution. Dr. Dantzig proved that in the 1940s and 1950s.

We note that with the simplex method, most errors and misformulations end up with the error "Cannot find an initial basis—execution terminated," or, worse than that, it finds an answer to a misformulated problem. (The former situation would occur, for example, when you have a stairstep supply curve that never gets out as far as a posited inelastic demand curve.) This is among the least helpful of error messages. It gives no guidance as to what is wrong in the linear programming model. Linear programming solvers are not self-debugging. This is in many ways a **less** useful error message than watching an iterative model fail to converge. With linear programming, the door is barred, and you cannot get in! That is all you know.

"Doesn't linear programming get to a unique solution?" Every economist and operations researcher worth his or her salt knows there is **there is no guarantee of uniqueness in complementarity, linear programming, or network equilibrium**. None at all. Period. With

linear programming, the cost hyperplane that is being minimized can quite easily be parallel with the plane of a facet of the simplex and thereby at the optimum touches the entire facet and thereby a number of vertices and all points in between. In short, the cost hyperplane can lie right along an entire facet of the simplex. In that situation, the linear programming answer (the quantity) is not unique. In such case, the quantity determined by the linear programming is in fact indeterminate. (This can be true for the primal and/or the dual.) Nonuniqueness is every bit as significant and prevalent with linear programming as with other methods.

This is actually a particularly thorny problem for linear programming. Many phenomena in economics lead to zero arbitrage prices. For example, in the presence of storage, natural gas seasonal price temporal variations arrive at the situation in which the storage operator is indifferent regarding which actual month he or she will inject to storage. That is precisely the situation in which the profit maximizing hyperplane falls exactly on a facet of his linear programming problem. The situation in which the solution is nonunique is not the exception in economics; it is the standard situation. It is the rule. It is the situation toward which the market gravitates. The multiple solution situation is the rule, not the exception.

The sufficiency conditions for uniqueness of a solution are fairly restrictive (and abstract). Think for a moment about the sufficiency conditions for a global optimum of a complex constrained optimization problem. One has a difficult (perhaps impossible) time determining whether a large, distributed, real world model with a complex optimand and hundreds of thousands of complex constraints meets such sufficiency conditions. The same is true with producer profit maximization equations. Little useful progress has been made on the issue of uniqueness of solution. In practical problems, algorithms tend empirically to find the same solution irrespective of initialization, but that is not a definitive proof.

"Complementarity solution algorithms are reliable. They are like linear programming, aren't they? They always get to a solution, don't they?" Such a conjecture could not be further from the truth, absolutely wrong. Gabriel op. cit. are very clear that complementarity solution algorithms that contain nonlinear relationships are multivariate Newton's methods or variants. They use gigantic Jacobean matrices or equivalent! **<u>They are iterative! Numerically iterative!</u>** They are not countably discrete descent algorithms such as the simplex method. They are numerically iterative, no different in concept from other numerically iterative algorithms such as fixed point, cobweb, Newton, secant, or other algorithms. There is no inherent algorithmic advantage to a complementarity problem, not at all. They are just as iterative and just as numerical (and substantially larger for the same problem) than network equilibrium algorithms. Any conjecture that somehow complementarity algorithms with nonlinearity in them are "safe" algorithms is absolutely incorrect on its face.

Copied directly from the dissertation of Stephen P. Dirkse, [26] Figure 11 states the definition of a Newton's method used to solve the complementarity problem. There is no question the solution method is **<u>iterative</u>**. There is no question that **<u>it requires the Jacobean matrix F(.) and its inverse</u>** iteratively. It is most definitely not a discrete descent method such as linear programming. It

[26] Dirkse, Stephen P., **<u>Robust Solution of Mixed Complementarity Problems</u>**, Ph.D. Dissertation, University of Wisconsin—Madison, 1994.

converges to one of the answers, falls into a local "hole," a local minimum, or veers off into outer space based on the conditioning of the Jacobean matrix. There is no guarantee it will find a solution even if one exists. It could iterate forever and diverge. There are sophisticated relaxation and other techniques required to, hopefully, get Newton's method to find a solution, but they do not work all the time. Anyone who argues that complementarity solution methods find solutions every time is just plain wrong. Newton's methods are no different from every other iterative algorithm—there is no guarantee of finding a solution even if there is one.

Figure 11: Dirkse's Definition of a Newton Method

Newton's method for solving the equation

$$F(x) = 0 \tag{NLE}$$

consists of two steps, approximation and zero-finding, applied repeatedly to produce a sequence of iterates $\{x^k\}$. In an approximation step, the function F is approximated, or linearized, at the point x^k by the affine function $A_k(\cdot)$ defined by

$$A_k(x) := F(x^k) + F'(x^k)(x - x^k). \tag{1.6}$$

The Newton point x_N^k is a zero of the approximation A_k, i.e. $A_k(x_N^k) = 0$. If the Jacobian matrix $F'(x^k)$ is nonsingular, this zero is unique, and is conceptually easy to find. Upon solving the matrix equation $F'(x^k)d^k = -F(x^k)$, the Newton point is given by $x_N^k = x^k + d^k$,

The Newton iteration is obtained by solving Dirkse's equation (1.6) for x when $A_k(x)=0$:

$$0 = F\left(x^k\right) + F'\left(x^k\right)\left(x - x^k\right)$$
$$\Rightarrow x^{k+1} = x^k - \left[F'\left(x^k\right)\right]^{-1} F\left(x^k\right)$$

Another major problem with Newton's methods in practice is that they are "locally convergent." When one gets very near to a solution, Newton's method (which depends via the Jacobean on a linear approximation to the nonlinear equation being solved) converges. (When one gets near to the answer, the linear approximation to the function gets better!) However, when one is not near an answer, the Jacobean can be illy conditioned, and the iteration can spin wildly out of control and fail to find any answer even if one exists, or as Gabriel op. cit. points out it can fall into a "local hole" that is not optimum. We have seen this many times with even the simplest of Newton iterations. When the literature says Newton's methods are "locally convergent," it **means** precisely that. Newton's methods have known and generally understood difficulties in non-local regions, which most regions of operation are. (Perhaps complementarians should adopt complicated genetic algorithms to maximize the probability of finding true global solutions.)

Using network equilibrium (implemented within the previous incarnation MarketBuilder), we have solved network microeconomic equilibrium problems with 25 million equations and unknowns (namely our multiregional North American Electricity Model, which had 3,500 nodes times 30 forward time points times 120 time intervals within each year times 2—price and quantity.) To solve a problem of that size using complementarity, one would need a Jacobean matrix with 25 million rows and 25 million columns. It would be a tremendously sparse matrix, and yet the precise position and magnitude of every entry in the sparse matrix would be crucial. One would need to recalculate and invert this 25 million by 25 million matrix every iteration in order to execute a Newton's method calculation to solve the problem. The Jacobean matrix changes every iteration. Dirkse's discussion of Newton iterations makes that clear. And the inverse is not sparse even though the matrix is sparse. There would be myriad matrix recalculations and solutions in order to (hopefully) solve the problem. It is irrelevant how much computer power one has, one is not going to solve that kind of a problem with a Newton complementarity approach. The matrix inversion "fuzz" (roundoff/truncation error) inverting a matrix that size alone is destined to doom the method.

We know that multivariate Newton's methods are every bit as iterative as a cobweb, and they can spiral out into space and not converge if the Jacobean matrices are illy conditioned or if matrix inversion (or linear system) roundoff and truncation error propagates. They are multidimensional, not single dimensional, so they go out of control and become poorly conditioned particularly when they are not local to a solution. There is no intrinsic reliability advantage between a multivariate Newton's method (which is in no way is guaranteed to solve a nonlinear complementarity problem even if a solution exists) and a simple fixed point cobweb or auction process. Quite the contrary, the simple, local fixed point cobweb or auction process is far MORE reliable, not less reliable. All these old, hackneyed arguments about solution reliability may have some validity for linear programming (which solves the wrong problem), but they are absolutely not true for complementarity. The EIA will be taking on a less, not more, reliable method if EIA deploys complementarity. The reality is that complementarity is four or more times too large in our example with a far less reliable, not more reliable, multidimensional Newton's method with a monster Jacobean formation and inversion required at every iteration to solve them. Plus you have to program them a command at a time. Who at EIA wants to program a 25 million row by 25 million column Jacobean matrix? The very idea of programming such a construct one command at a time in FORTRAN looking GAMS or AMPL code defies reality. That simply cannot meet the need of an organization such as the EIA. It is labor intensive and error prone.

The author has a good deal of experience with multivariate Newton's methods in this and other contexts, particularly maximum likelihood contexts. He has discovered (the hard way) that these algorithms are highly local in nature and difficult to converge. Highly, highly local. If you start them at random or start them some distance away from the optimum answer, they spiral off into space and do not converge, or they fall into a local hole. The literature tells us they are locally convergent, not globally convergent. The literature tells us that they can get trapped in a local hole and fail to migrate out and get to the correct answer, i.e., they can fall into a hole and hang up at a non-solution.

Finally, we have no idea why anyone would ever formulate a linear complementarity problem (except for research purposes). You already have linear programming to solve linear problems. The only motivation to formulate a complementarity problem is if it is a nonlinear problem, either a nonlinear objective function or nonlinear constraints or both. If complementarity turns out to be a breakthrough, it will be a breakthrough for nonlinear problems. We have plenty of ways to solve linear problems already. Dr. Dantzig told us how to pivot sixty years ago. Not much new there. It is the mixed and fully nonlinear problems that we need to solve, and the nonlinear complementarity formulation is simply too big.

4.6.2 Analytical Substitution Trumps Numerical Algorithm

The notion, proven and demonstrated in this report, that you can substitute within complementarity equations **analytically** and end up with a much smaller and easier network equilibrium, is a phenomenal notion. Such substitution eliminates the need for complementarity equations ever having to be solved numerically; they can all be solved analytically. The analytically reduced equation set is much easier to solve. The complementarity people are happy—they get their complementarity formulation! It is solved analytically. The subsequent substitution reduces the complementarity equations from a "hubs plus activities by hubs plus activities" size to a "hubs only" size, a massive reduction in dimensionality. If the complementarity problem has a solution, then certainly the network equilibrium model you get after the substitutions gives the same answer. The network equilibrium solution is much more likely to find that answer and not be prone to algorithm, roundoff, truncation, conditioning, massive size, or other troublesome errors that face multivariate Newton's methods.

4.6.3 Complementarity Solution Algorithms

From Gabriel op. cit., page 377:

Most methods to solve the NCP [Nonlinear Complementarity Problem] *involve a Newton approach, but the methods differ in the type of nonsmooth equation and merit function. They also differ in the way in which a direction is chosen for moving to the next iterate, and these differences are often due to differences in the type of nonsmooth equation and merit function. Most of the methods can be understood to be descent algorithms, i.e., they work to minimize a merit function. However, since the merit function often has local minima with a positive value (where the nonsmooth equation is not satisfied), these methods can get stuck at local minima.*

Does this sound like a reliable algorithm guaranteed to find the minimum of a complicated (and as we will show oversized) merit function? Not a chance. Newton's methods are among the most difficult and unstable iterative algorithms, very amenable to spinning off into space or getting stuck at local non-solutions. This has been well known in the literature for years. Gabriel op. cit. has articulated it in their monograph. None of the algorithms he presents or that we have found in the literature guarantees a solution. They are fundamentally iterative. Thankfully, we have proven that they can be solved analytically and thereby avoided.

As Gabriel op. cit. has stated, the most promising nonlinear complementarity algorithms are Newton's methods descended from merit functions. The most illustrative are those descended from the Fisher-Burmeister type of function. (It seems that Fischer-Burmeister types of functions, analytical tricks, stimulated interest in direct solutions to Kuhn-Tucker conditions written as complementarity equations.) Consider the Fischer-Burmeister function

$$\phi(a,b) = \left(a^2 + b^2\right)^{\frac{1}{2}} - (a+b)$$

This function will take the value zero when

$$a^2 + b^2 = (a+b)^2 \Rightarrow 2ab = 0$$

This implies two conditions

$$\begin{aligned} b \neq 0 \Rightarrow a = 0 \\ a \neq 0 \Rightarrow b = 0 \end{aligned}$$

If we substitute the first condition into the square root equation above, we obtain

$$\phi(0,b) = 0 = \left(b^2\right)^{\frac{1}{2}} - b = |b| - b \Rightarrow |b| = b \Rightarrow b > 0$$

If we substitute the second condition into the square root equation above, we obtain

$$\phi(a,0) = 0 = \left(a^2\right)^{\frac{1}{2}} - a = |a| - a \Rightarrow |a| = a \Rightarrow a > 0$$

Thus, the Fischer Burmeister function gives the result

$$\begin{aligned} b > 0 \Rightarrow a = 0 \\ a > 0 \Rightarrow b = 0 \end{aligned}$$

Thus, the Fischer Burmeister type of function has the property that

$$\phi(a,b) = 0 \Leftrightarrow a \geq 0,\ b \geq 0,\ ab = 0$$

These are precisely the complementarity conditions, which appear exactly as if they were Kuhn-Tucker Conditions. When the Fischer-Burmeister function has a zero value, the complementarity conditions are satisfied. The Fischer-Burmeister function is continuously differentiable everywhere but at the origin. What if we took every complementarity equation and wrote it in Fisher-Burmeister function form? What would the Newton iterations look like? How difficult would they be?

4.6.4 Apply the Fischer-Burmeister Equations to the Complementarity Problem

It is instructive to lay out the Fischer-Burmeister formulation of the complementarity solution to see how huge, ponderous, and unreliable the system becomes. We are going to represent the equations in detail so that we can observe the magnitude of programming that is required.

4.6.4.1 Four Demand Equations (4 Total)

$$\phi\left(\left[\lambda_{A,17}-p_{17}\left(q_{A,17}\right)\right],q_{A,17}\right)=0=\left\{\left[\lambda_{A,17}-p_{17}\left(q_{A,17}\right)\right]^{2}+q_{A,17}{}^{2}\right\}^{\frac{1}{2}}-\left\{\left[\lambda_{A,17}-p_{17}\left(q_{A,17}\right)\right]+q_{A,17}\right\}$$

$$\phi\left(\left[\lambda_{B,18}-p_{18}\left(q_{B,18}\right)\right],q_{B,18}\right)=0=\left\{\left[\lambda_{B,18}-p_{18}\left(q_{B,18}\right)\right]^{2}+q_{B,18}{}^{2}\right\}^{\frac{1}{2}}-\left(\left[\lambda_{B,18}-p_{18}\left(q_{B,18}\right)\right]+q_{B,18}\right)$$

$$\phi\left(\left[\lambda_{C,19}-p_{19}\left(q_{C,19}\right)\right],q_{C,19}\right)=0=\left\{\left[\lambda_{C,19}-p_{19}\left(q_{C,19}\right)\right]^{2}+q_{C,19}{}^{2}\right\}^{\frac{1}{2}}-\left(\left[\lambda_{C,19}-p_{19}\left(q_{C,19}\right)\right]+q_{C,19}\right)$$

$$\phi\left(\left[\lambda_{D,20}-p_{20}\left(q_{D,20}\right)\right],q_{D,20}\right)=0=\left(\left[\lambda_{D,20}-p_{20}\left(q_{D,20}\right)\right]^{2}+q_{D,20}{}^{2}\right)^{\frac{1}{2}}-\left(\left[\lambda_{D,20}-p_{20}\left(q_{D,20}\right)\right]+q_{D,20}\right)$$

4.6.4.2 Four Supply Equations (4 Total)

$$\phi\left(\left[MC_{21}\left(q_{21,E}\right)-\lambda_{21,E}\right],q_{21,E}\right)=0=\left\{\left[MC_{21}\left(q_{21,E}\right)-\lambda_{21,E}\right]^{2}+q_{21,E}{}^{2}\right\}^{\frac{1}{2}}-\left(\left[MC_{21}\left(q_{21,E}\right)-\lambda_{21,E}\right]+q_{21,E}\right)$$

$$\phi\left(\left[MC_{22}\left(q_{22,F}\right)-\lambda_{22,F}\right],q_{22,F}\right)=0=\left(\left[MC_{22}\left(q_{22,F}\right)-\lambda_{22,F}\right]^{2}+q_{22,F}{}^{2}\right)^{\frac{1}{2}}-\left(\left[MC_{22}\left(q_{22,F}\right)-\lambda_{22,F}\right]+q_{22,F}\right)$$

$$\phi\left(\left[MC_{23}\left(q_{23,G}\right)-\lambda_{23,G}\right],q_{23,G}\right)=0=\left(\left[MC_{23}\left(q_{23,G}\right)-\lambda_{23,G}\right]^{2}+q_{23,G}{}^{2}\right)^{\frac{1}{2}}-\left(\left[MC_{23}\left(q_{23,G}\right)-\lambda_{23,G}\right]+q_{23,G}\right)$$

$$\phi\left(\left[MC_{24}\left(q_{24,H}\right)-\lambda_{24,H}\right],q_{24,H}\right)=0=\left(\left[MC_{24}\left(q_{24,H}\right)-\lambda_{24,H}\right]^{2}+q_{24,H}{}^{2}\right)^{\frac{1}{2}}-\left(\left[MC_{24}\left(q_{24,H}\right)-\lambda_{24,H}\right]+q_{24,H}\right)$$

4.6.4.3 Sixteen Transportation Output Relationships (16 Total)

$$\phi\left(\left[VOC_{1}+\lambda_{21,E}g_{1}'\left(q_{1,A}\right)-\lambda_{A,17}\right],q_{1,A}\right)=$$
$$0=\left\{\left[VOC_{1}+\lambda_{21,E}g_{1}'\left(q_{1,A}\right)-\lambda_{A,17}\right]^{2}+q_{1,A}{}^{2}\right\}^{\frac{1}{2}}-\left(\left[VOC_{1}+\lambda_{21,E}g_{1}'\left(q_{1,A}\right)-\lambda_{A,17}\right]+q_{1,A}\right)$$

$$\phi\left(\left[VOC_{2}+\lambda_{22,F}g_{2}'\left(q_{2,A}\right)-\lambda_{A,17}\right],q_{2,A}\right)=$$
$$0=\left(\left[VOC_{2}+\lambda_{22,F}g_{2}'\left(q_{2,A}\right)-\lambda_{A,17}\right]^{2}+q_{2,A}{}^{2}\right)^{\frac{1}{2}}-\left(\left[VOC_{2}+\lambda_{22,F}g_{2}'\left(q_{2,A}\right)-\lambda_{A,17}\right]+q_{2,A}\right)$$

$$\phi\left(\left[VOC_{3}+\lambda_{23,G}g_{3}'\left(q_{3,A}\right)-\lambda_{A,17}\right],q_{3,A}\right)=$$
$$0=\left(\left[VOC_{3}+\lambda_{23,G}g_{3}'\left(q_{3,A}\right)-\lambda_{A,17}\right]^{2}+q_{3,A}{}^{2}\right)^{\frac{1}{2}}-\left(\left[VOC_{3}+\lambda_{23,G}g_{3}'\left(q_{3,A}\right)-\lambda_{A,17}\right]+q_{3,A}\right)$$

$$\phi\left(\left[VOC_4+\lambda_{24,H}g_4'\left(q_{4,A}\right)-\lambda_{A,17}\right],q_{4,A}\right)=$$

$$0=\left(\left[VOC_4+\lambda_{24,H}g_4'\left(q_{4,A}\right)-\lambda_{A,17}\right]^2+q_{4,A}{}^2\right)^{\frac{1}{2}}-\left(\left[VOC_4+\lambda_{24,H}g_4'\left(q_{4,A}\right)-\lambda_{A,17}\right]+q_{4,A}\right)$$

$$\phi\left(\left[VOC_5+\lambda_{21,E}g_5'\left(q_{5,B}\right)-\lambda_{B,18}\right],q_{5,B}\right)=$$

$$0=\left(\left[VOC_5+\lambda_{21,E}g_5'\left(q_{5,B}\right)-\lambda_{B,18}\right]^2+q_{5,B}{}^2\right)^{\frac{1}{2}}-\left(\left[VOC_5+\lambda_{21,E}g_5'\left(q_{5,B}\right)-\lambda_{B,18}\right]+q_{5,B}\right)$$

$$\phi\left(\left[VOC_6+\lambda_{22,F}g_6'\left(q_{6,B}\right)-\lambda_{B,18}\right],q_{6,B}\right)=$$

$$0=\left(\left[VOC_6+\lambda_{22,F}g_6'\left(q_{6,B}\right)-\lambda_{B,18}\right]^2+q_{6,B}{}^2\right)^{\frac{1}{2}}-\left(\left[VOC_6+\lambda_{22,F}g_6'\left(q_{6,B}\right)-\lambda_{B,18}\right]+q_{6,B}\right)$$

$$\phi\left(\left[VOC_7+\lambda_{23,G}g_7'\left(q_{7B}\right)-\lambda_{B,18}\right],q_{7B}\right)=$$

$$0=\left(\left[VOC_7+\lambda_{23,G}g_7'\left(q_{7B}\right)-\lambda_{B,18}\right]^2+q_{7B}{}^2\right)^{\frac{1}{2}}-\left(\left[VOC_7+\lambda_{23,G}g_7'\left(q_{7B}\right)-\lambda_{B,18}\right]+q_{7B}\right)$$

$$\phi\left(\left[VOC_8+\lambda_{24,H}g_8'\left(q_{8,B}\right)-\lambda_{B,18}\right],q_{8,B}\right)=$$

$$0=\left(\left[VOC_8+\lambda_{24,H}g_8'\left(q_{8,B}\right)-\lambda_{B,18}\right]^2+q_{8,B}{}^2\right)^{\frac{1}{2}}-\left(\left[VOC_8+\lambda_{24,H}g_8'\left(q_{8,B}\right)-\lambda_{B,18}\right]+q_{8,B}\right)$$

$$\phi\left(\left[VOC_9+\lambda_{21,E}g_9'\left(q_{9,C}\right)-\lambda_{C,19}\right],q_{9,C}\right)=$$

$$0=\left(\left[VOC_9+\lambda_{21,E}g_9'\left(q_{9,C}\right)-\lambda_{C,19}\right]^2+q_{9,C}{}^2\right)^{\frac{1}{2}}-\left(\left[VOC_9+\lambda_{21,E}g_9'\left(q_{9,C}\right)-\lambda_{C,19}\right]+q_{9,C}\right)$$

$$\phi\left(\left[VOC_{10}+\lambda_{22,F}g_{10}'\left(q_{10,C}\right)-\lambda_{C,19}\right],q_{10,C}\right)=$$

$$0=\left(\left[VOC_{10}+\lambda_{22,F}g_{10}'\left(q_{10,C}\right)-\lambda_{C,19}\right]^2+q_{10,C}{}^2\right)^{\frac{1}{2}}-\left(\left[VOC_{10}+\lambda_{22,F}g_{10}'\left(q_{10,C}\right)-\lambda_{C,19}\right]+q_{10,C}\right)$$

$$\phi\left(\left[VOC_{11}+\lambda_{23,G}g_{11}'\left(q_{11,C}\right)-\lambda_{C,19}\right],q_{11,C}\right)=$$

$$0=\left(\left[VOC_{11}+\lambda_{23,G}g_{11}'\left(q_{11,C}\right)-\lambda_{C,19}\right]^2+q_{11,C}{}^2\right)^{\frac{1}{2}}-\left(\left[VOC_{11}+\lambda_{23,G}g_{11}'\left(q_{11,C}\right)-\lambda_{C,19}\right]+q_{11,C}\right)$$

$$\phi\left(\left[VOC_{12}+\lambda_{24,H}g_{12}'\left(q_{12,C}\right)-\lambda_{C,19}\right],q_{12,C}\right)=$$

$$0=\left(\left[VOC_{12}+\lambda_{24,H}g_{12}'\left(q_{12,C}\right)-\lambda_{C,19}\right]^2+q_{12,C}{}^2\right)^{\frac{1}{2}}-\left(\left[VOC_{12}+\lambda_{24,H}g_{12}'\left(q_{12,C}\right)-\lambda_{C,19}\right]+q_{12,C}\right)$$

$$\phi\left(\left[VOC_{13}+\lambda_{21,E}g_{13}'\left(q_{13,D}\right)-\lambda_{D,20}\right],q_{13,D}\right)=$$
$$0=\left(\left[VOC_{13}+\lambda_{21,E}g_{13}'\left(q_{13,D}\right)-\lambda_{D,20}\right]^{2}+q_{13,D}{}^{2}\right)^{\frac{1}{2}}-\left(\left[VOC_{13}+\lambda_{21,E}g_{13}'\left(q_{13,D}\right)-\lambda_{D,20}\right]+q_{13,D}\right)$$

$$\phi\left(\left[VOC_{14}+\lambda_{22,F}g_{14}'\left(q_{14,D}\right)-\lambda_{D,20}\right],q_{14,D}\right)=$$
$$0=\left(\left[VOC_{14}+\lambda_{22,F}g_{14}'\left(q_{14,D}\right)-\lambda_{D,20}\right]^{2}+q_{14,D}{}^{2}\right)^{\frac{1}{2}}-\left(\left[VOC_{14}+\lambda_{22,F}g_{14}'\left(q_{14,D}\right)-\lambda_{D,20}\right]+q_{14,D}\right)$$

$$\phi\left(\left[VOC_{15}+\lambda_{23,G}g_{15}'\left(q_{15,D}\right)-\lambda_{D,20}\right],q_{15,D}\right)=$$
$$0=\left(\left[VOC_{15}+\lambda_{23,G}g_{15}'\left(q_{15,D}\right)-\lambda_{D,20}\right]^{2}+q_{15,D}{}^{2}\right)^{\frac{1}{2}}-\left(\left[VOC_{15}+\lambda_{23,G}g_{15}'\left(q_{15,D}\right)-\lambda_{D,20}\right]+q_{15,D}\right)$$

$$\phi\left(\left[VOC_{16}+\lambda_{24,H}g_{16}'\left(q_{16,D}\right)-\lambda_{D,20}\right],q_{16,D}\right)=$$
$$0=\left(\left[VOC_{16}+\lambda_{24,H}g_{16}'\left(q_{16,D}\right)-\lambda_{D,20}\right]^{2}+q_{16,D}{}^{2}\right)^{\frac{1}{2}}-\left(\left[VOC_{16}+\lambda_{24,H}g_{16}'\left(q_{16,D}\right)-\lambda_{D,20}\right]+q_{16,D}\right)$$

4.6.4.4 Eight Balance Equations

$$q_{A,17}-q_{1,A}-q_{2,A}-q_{3,A}-q_{4,A}=0$$
$$q_{B,18}-q_{5,B}-q_{6,B}-q_{7,B}-q_{8,B}=0$$
$$q_{C,19}-q_{9,C}-q_{10,C}-q_{11,C}-q_{12,C}=0$$
$$q_{D,20}-q_{13,D}-q_{14,D}-q_{15,D}-q_{16,D}=0$$
$$g_{1}\left(q_{1,A}\right)+g_{5}\left(q_{5,B}\right)+g_{9}\left(q_{9,C}\right)+g_{13}\left(q_{13,D}\right)-q_{21,E}=0$$
$$g_{2}\left(q_{2,A}\right)+g_{6}\left(q_{6,B}\right)+g_{10}\left(q_{10,C}\right)+g_{14}\left(q_{14,D}\right)-q_{22,F}=0$$
$$g_{3}\left(q_{3,A}\right)+g_{7}\left(q_{7,B}\right)+g_{11}\left(q_{11,C}\right)+g_{15}\left(q_{15,D}\right)-q_{23,G}=0$$
$$g_{4}\left(q_{4,A}\right)+g_{8}\left(q_{8,B}\right)+g_{12}\left(q_{12,C}\right)+g_{16}\left(q_{16,D}\right)-q_{24,H}=0$$

In this situation, we have a set of 32 equations and 32 unknowns. This would require a 32 by 32 Jacobean matrix, and that Jacobean matrix will be very sparse. (Each of the 32 equations does not contain all 32 variables. That is what renders the Jacobian matrix so sparse and difficult to implement and manage.) Furthermore, as indicated by Gabriel op. cit., the algorithm is just as likely to “fall into a hole” at a local minimum and never get to the global minimum. Furthermore, it is difficult to impossible to understand the convexity of the constraints in general.

Why the number 32? There are 8 prices, and there are 24 quantities. The reality of Newton iteration is this: **The size of the Jacobean matrix (which must be refreshed and inverted on every Newton iteration) is the number of quantities PLUS the number of prices times the number**

of quantities PLUS the number of prices. This is such a mammoth size matrix for real world problems that it simply will not allow sufficient disaggregation of the world transportation system. When one considers temporality in the problem, the size explodes by the expanded "Cartesian product" that now considers the number of forward time points in both rows and columns. The consequent overall size limitation of the Jacobean matrix squeezes down the number of transportation nodes that can be effectively implemented. Complementarity will squeeze the actual problem size to be too small to meet EIA's need to represent a world transportation matrix in sufficient detail. EIA is likely to suffer from extreme aggregation, and if that occurs its models would not be suitable if implemented in a complementarity framework.

4.7 Summary—What Do We Learn from the Complementarity Formulation?

What have we learned? There are a few salient things:

1. We now understand the essence of the complementarity formulation in the specific context of a transportation example that we have posited. **The complementarity formulation is innately derived from monolithic global welfare maximization for a competitive market.** Gabriel op. cit. specifically recognize that, and they recognize it explicitly and systematically. The complementarity problem is derived from the Kuhn-Tucker conditions for monolithic global welfare maximization. This is the approach Samuelson first put forth in 1952; it is not "new tech" or "high tech." It and complementarity have been known for six decades. There is absolutely nothing new about complementarity. It is just Kuhn-Tucker conditions rewritten.

2. The monolithic global welfare maximization problem **naturally decomposes into two pieces**: (1) a cost-minimizing network transportation problem and (2) a supply and demand problem derived from monolithic global welfare maximization. Samuelson observed this in 1952, and yet it seems to have been lost in the annals of time and marketing by modelers since then. Whether or not the problem is linear, one can isolate the transportation-only problem and solve it separately but interconnectedly with the supply and demand problem. This means that EIA can use the same or an alternative solution to the transportation problem (other than complementarity) and interface it with complementarity representations of supply and demand. EIA need not adopt and use a single methodology.

3. The **computer architecture necessary to solve the complementarity problem is ominous and difficult**. Gabriel op. cit. have discussed a number of complementarity algorithms in their monograph, and they show some modern GAMS code (AMPL appears to be similar). The complementarity solution algorithms are all "full rank," spanning the sum of all the Lagrange multipliers plus all the quantities times the sum of all the Lagrange multipliers plus all the quantities. (In the linear programming analogy, they span the sum of the rows plus columns of the constraint matrix times that same sum.) It is a "simultaneous solution" method that represents the problem in a large and ponderous fashion. There is no way anyone has published to systemize the architecture in order to allow expansion, contraction, or modification of the underlying transportation grid. If one observes in Gabriel op. cit. the examples of GAMS or AMPL input code, one should be taken completely aback. One knows that you have to write that code line by line and input that data using cumbersome line by line commands to implement or modify a model using literal 1970s vintage computer technology. You have to

"deliver things to a solver." Within GAMS or AMPL, modeling is **computer coding**. Modeling via computer coding is not viable for operational models or for organizations such as EIA, who have to model transportation systems rapidly and accurately. There is much more to come on that pivotal point.

4. **Complementarity requires you to "write equations."** EIA does not want to write equations when you add, delete, build, or modify a model. In fact, the last thing you want to be doing is programming or writing or structuring equations or manipulating constraint matrixes for linear programs. To be compelled to structure or write equations is tantamount to expanding your model turnaround time by two orders of magnitude.

5. Solution algorithms for complementarity are **large, concentrated, monolithic algorithms**, the equivalent of a big command line oriented, GAMS-type or AMPL-type subroutine. They have a dependency matrix that is the sum of all the hubs in the model (associated with "rows") plus all the economic activities in the model (associated with "columns") times itself, a very large matrix. Complementarity does nothing to alleviate the implementation difficulties of linear programming formulations, which are known to be burdensome at best. It in fact exacerbates those difficulties measurably. Complementarity is very much larger than linear programming because of the generalization that supplies, transportation links, and demands can be represented nonlinearly rather than linearly and because of the much larger size of the Jacobean than the constraint matrix. Complementarity will be even more computationally difficult than linear programming because of the size and sparse structure of the Jacobean matrix, which is key to the solution.

6. There is absolutely **no effort in complementarity given to inserting analytic solutions** or "sub-solutions" to subsets or all of the Kuhn-Tucker or complementarity equations. Analytic solutions to supplier problems, transportation problems, or demand problems are absent from complementarity (there is no attention given to that in Gabriel op. cit. for example or any other papers of which we are aware). And yet direct analytic solution of a subset of the Kuhn-Tucker equations is **far** preferable to brute force numerical solution. If you have an analytic solution, an analytic function that solves the problem, you should always use it. Unconditionally. Period. Analytic solutions are always correct, and they consume much less execution time. Analytical solutions allow much smaller solutions. They are auditable and reviewable by modelers and professionals. Analytic solutions contain insight in their own right. Convergence becomes a non-sequitur because they are a single analytic function rather than a complex (and unnecessary) numerical procedure. The function can be analyzed, not merely run through a "solver" without knowing the function. Importantly, they save and free up the dimensionality and computational time that would otherwise be consumed to solve them over and over. Why would one solve a transporter optimization problem in a complementarity procedure when one could solve it once analytically and input that analytic solution directly? Microeconomic theory is predicated on analytical solutions to elemental problems and network interconnection of those analytical solutions in a way that will be clear in the subsequent section. Microeconomic theory is not predicated on the brute force simultaneous numerical solution using a centralized "solver" or algorithm. We continue to be surprised that no one has done what is done later in this paper (as far as we know)—**derive the analytic solution to the complementarity**

<u>problem and thereby eliminate the need for the complementarity problem altogether</u>. Analytics will trump numerics.

7. **<u>Complementarity is not suited to large economic problems such as transportation.</u>** We can see from the complexity of the complementarity problem, even for the simple case discussed in the previous section, that the complementarity solution doesn't really fit the transportation problem. It is too bulky. It is much more productive to adopt a technique that is specifically network oriented. After all, transportation is a network-oriented problem.

8. **<u>Complementarity is an operations research technique</u>**. The Gabriel op. cit. book is an operations research text. It is not an economics text. Complementarity is not an economically derived approach to the problem. As such, whether or not it may be technically correct or consistent with solid microeconomic theory, it is so large and unworkable that EIA should avoid using it for spatial network problems. There are better, more efficient ways, especially the one in Section 5.

9. **<u>If complementarity is as good as advocated, there is no reason ever to use linear programming again</u>**. Linear programming is a limiting special case of complementarity, which is derived from nonlinear optimization. Nonlinear optimization and complementarity are more general. Consider activities such as individual refinery modeling, long done by linear programming. It would be far more compact, accurate, and direct simply to replace the linear objective and linear constraints with nonlinear forms and solve it with Kuhn-Tucker conditions reduced to complementarity problems. The real advantage of complementarity is that it might eliminate the need for linear programming in contexts other than economic problems. If complementarity proves to be as good as advertised, there would never again be justification to use linear programming. Complementarity would trump it. Linear programming would proceed into the annals of history with things like hitchin' posts, Newtonian physics, APL programming, or bloodletting to cure plague and other ill humors.

10. **<u>Complementarity is not explainable to management or constituents.</u>** This section has been dense and difficult. We conjecture that few senior managers will be reading this. Complementarity and its motivation or correctness simply cannot be explained to the boss or the decision maker. It is not a natural and direct agent-based approach. It is indirect. How many times have we heard people say "Complementarity is the same as other approaches in economics? You can trust us on that." As we will see, network microeconomics is much more compelling and easily articulated, and the results, formulation, and rationale are much more easily communicated. In fact, the original version of network microeconomic modeling was developed specifically to be accessible by executives and senior managers as well as analysts.

Complementarity is a way to express and hopefully solve nonlinear monolithic optimization problems, particularly monolithic constrained optimization problems that are laden with constraints, devolved as complementarity is from Kuhn-Tucker conditions for such problems. Complementarity could well replace and substitute for linear programming for monolithic optimization problems outside the economic realm, obviating the need for linear programming altogether. We see potential benefits from complementarity for problems **<u>other than</u>** economic problems, and we perceive benefits from superseding and replacing linear programming altogether. (Of course there are other nonlinear optimization solution methods besides complementarity, and it is not clear when or whether any method has advantages over another.) As an example, if one simply applied a gradient projection method in the positive orthant to the original monolithic global welfare maximization problem in Section 4, that would solve the problem by construction. The local optimum thus obtained (which is what both the Kuhn-Tucker conditions as well as the gradient projection methods find) would solve the complementarity equations by construction because it solves the Kuhn-Tucker conditions. That is the intrinsic property of a local optimum no matter how one might find it, even if one never directly approaches the problem using the complementarity equations or a complementarity algorithm. This is not an insignificant point. One does not have to have a complementarity algorithm to solve a complementarity problem. If one solves an optimization problem and finds a local optimum using any method at all—gradient projection, barrier/penalty functions, self-scaling variable metric, genetic algorithms, etc.—he or she has de facto solved the complementarity equations. You to not have to solve the complementarity equations to solve the complementarity equations! The most important and cogent understanding one can communicate about complementarity equations is that they are merely restatements of Kuhn-Tucker conditions. The two are isomorphic and entirely equivalent. Network microeconomic equilibrium is a simpler, more straightforward way to solve complementarity equations when there is equivalence.

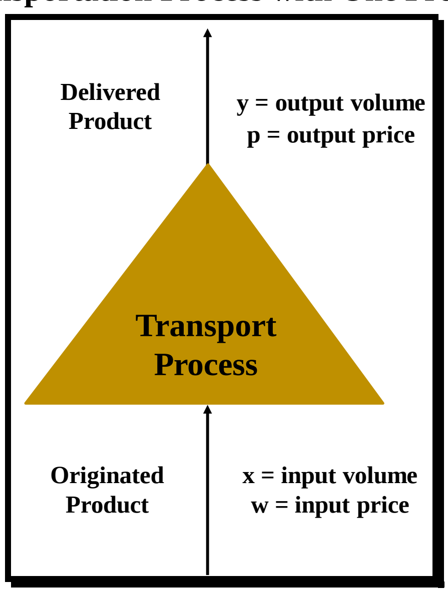
We will return to this theme in Section 7, showing that one can completely solve the complementarity equations by substitution, and when one does so, one simplifies and downsizes to the network equilibrium equations.

5 A GENERAL EQUILIBRIUM NETWORK FORMULATION OF THE TRANSPORTATION PROBLEM

Complementarity is not well suited for transportation or other spatial economic problems, as we have seen in the previous section and shall see in spades by comparison in this section and in Section 7. The direct network microeconomic approach is much simpler and more direct, and it does not impose the risks of a monolithic global optimization approach. This section approaches the spatial transportation problem from a fundamental microeconomic perspective, one that will yield a much simpler mathematical representation and a much simpler solution than the complex and oversized complementarity approach, which as we and Gabriel op. cit. have stated is derived (precisely as Samuelson first did in 1952 and others followed) from monolithic global welfare maximization. There is no argument about this—complementarity devolves fundamentally from monolithic global welfare maximization. Gabriel op. cit. themselves emphasize that salient point. Microeconomic theory, a direct approach derived from actions by individual agents, motivates us to consider an alternative, agent-based approach. There are cases in which the two are equivalent, and there are cases in which they are not. There are never cases in which complementarity is preferred.

Let us pose the transportation asset operation problem by positing a single profit-maximizing "producer," a "producer" who is a provider of transportation service. That transportation provider is suspended within and operates within a price system. We depict the transportation provider suspended in a price system graphically as in Figure 12. We conceive the transportation process as a technique that draws commodity from an origin point and delivers it to a terminus point, and the transporter faces a price at the origin point and the delivery point.

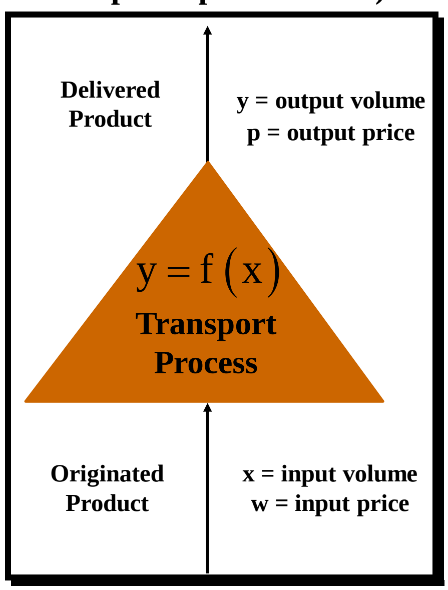
Figure 12: A Transportation Process with One Product and One Input



The input-output relationship for a transportation process is dictated, as are all economic activities everywhere throughout a spatial economic system, by an output-input function, equivalently a classical economic production function. A production function specifies the relationship between the quantity delivered at the destination of the transportation process and the quantity input at the origin of the transportation process. Such a production function takes account of such phenomena as compression losses or LNG boiloff and burnoff.[27]

We emphasize that the transportation process is possessed of a production function, i.e., a loss function, no differently from any and every other economic process that comprises an economy or an energy market. The emphasis on the production function is apparent in Figure 13. In an important sense, a transportation process is no more than a "long" energy conversion process. A pipeline converts gas in Louisiana to gas in Ohio. An LNG tanker converts LNG FOB Qatar to LNG FOB Zeebrugge, Belgium. An internal combustion gasoline powered automobile converts gasoline to vehicle miles. A furnace converts natural gas to warm air through the register. All are characterized by production functions in precisely the same way as we summarize in Figure 13.

Figure 13: A Transportation Process in Characterized by a Production Function (an Output-Input Function)



It is important to emphasize that this production function representation of a transportation process is completely general because the production function f() can be any function. In practice, it can be carefully crafted to represent the physics, chemistry, and combustion of the particular transportation process (as it would be for an LNG tanker, a crude tanker, or a large diameter pipe). It specifies, for every possible delivery level y, the quantity x that would have to be uptaken at the origination point. The transportation production function takes full account of such issues as steaming losses, compression losses, boiloff, capacity available, etc.

In the economic context, "production" in the sense of a transportation process means moving a fuel from one physical location to another physical location. The "input," the factor in the economic sense, is the fuel at the origin of the transportation process. The "output" is the same fuel at the terminus of the transportation process (but generally a smaller quantity due to burnoff or compression losses). There are losses and fixed and variable costs along the way, which we shall represent shortly, but the production function explicitly represents the losses (i.e., shrinkage)

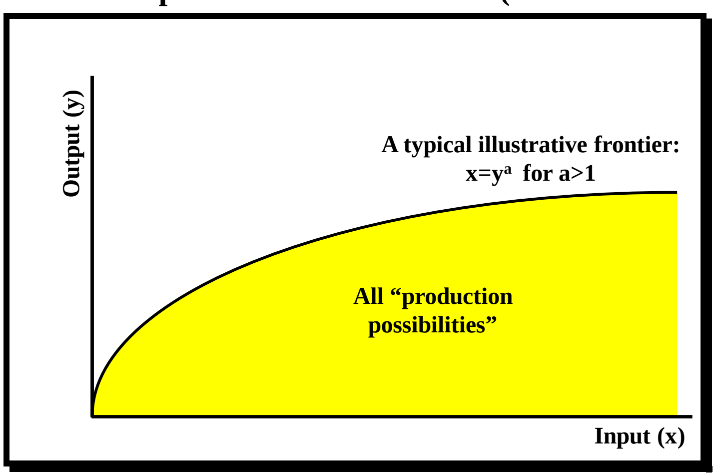
[27] Complementarity used a production function for every pipeline also. This isn't a unique or surprising representation.

in transportation. No one disagrees that there exists a production function that characterizes the transportation facility. Indeed, such an assumption is no different from what was assumed for the monolithic global welfare maximization approach, i.e., an input-output relationship for every transportation process.

In our experience, models and modelers tend to underemphasize that production is represented as a set of all possible outputs and inputs. Examples abound in the real world—"heat rates," "compression losses," "shrinkage," "input-output coefficients," "boil-off," and so forth. There is nothing magical or arcane in the specification of a production function. It occurs repeatedly throughout the energy system. Generators, pipes, refineries, and other energy processes very carefully measure their inputs, outputs, and losses. The previous complementarity development used the selfsame production functions in the form of input-output relationships.

So what do these production functions look like? What would the production possibility set for a pipeline or a tanker look like? We have presented in Figure 14 an example production possibility set, whose boundary is the production function.[28] We see in Figure 14 declining marginal efficiency of transportation as output gets to be larger in magnitude. This phenomenon tends to be ubiquitously observed. Ultimately transportation output will reach installed capacity, in which the marginal efficiency of transportation goes to zero. Even before one reaches that extreme, however, we would expect to observe that compression losses get higher the more fully you load transportation process, i.e., the more volume you attempt to move through it. (Your car is a great example. The faster you go, the higher your losses, i.e., the worse your mileage—the higher your fuel consumption.)

Figure 14: An Example Production Function (Production Possibility Set)



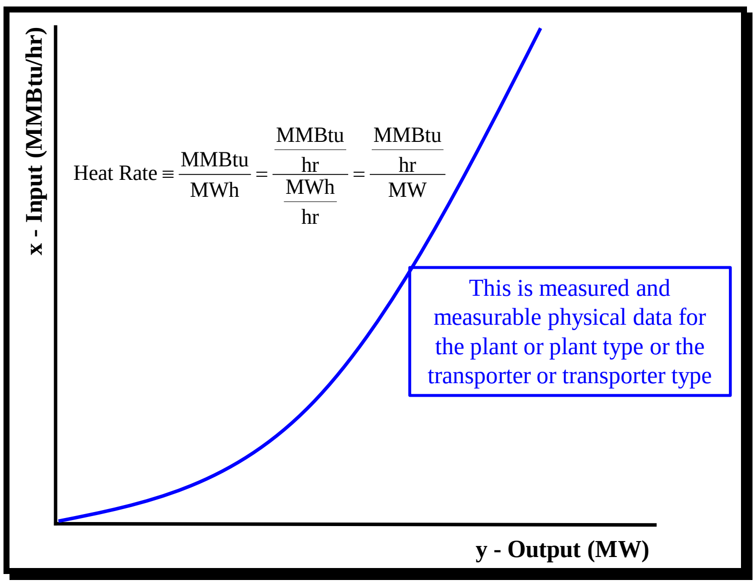
How would you quantify this production function for a plant or a transporter? How would you measure an input-output curve for a plant or a transporter? The answer to the question is not difficult. You would simply observe and measure various levels of inputs and outputs. This is done

[28] This specific production function is **not** the one that is resident within the ArrowHead model of transportation. This one is simpler than the one used there, yet it is quite illustrative of the producer profit maximization approach we use. ArrowHead uses an intertemporal quadratic form. The generality of this development is not limited by the specifics of the production function selected.

routinely, all the time, by equipment vendors. We see it time and again in the power industry, people measuring and reporting fuel consumed by a power plant and MWh output from the power plant. Vendors want to know how their equipment performs. Government and regulatory agencies ask for reports of uptake and delivery. Asset owners want to know where they are spending their fuel dollars (i.e., their input factor dollars).

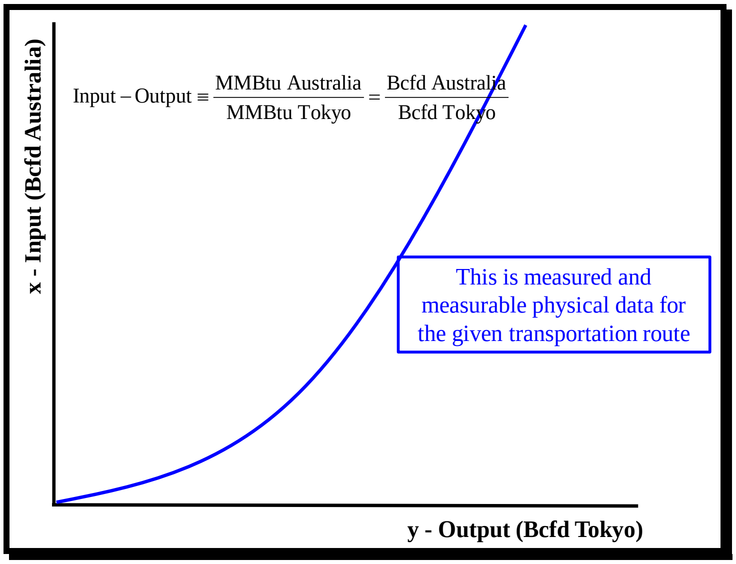
Even though a transportation process is not a power plant, it is useful to appeal to the power plant analogy of a "heat rate." A heat rate is the number of units of **input** you need (generally expressed in terms of Btus) to produce one unit of **output** (generally expressed in terms of MWh). You would for the transportation process involved, say an LNG tanker from the Northwest Shelf of Australia to Tokyo, Japan, plot the Btus of gas taken on board in Australia versus the Btus of gas delivered to the terminal in Tokyo, Japan. That would be the input-output curve for that LNG tanker route, and it would, of course, be specific to that route. It would have the generic shape for a power plant in Figure 15. Precisely the identical analogy would hold for the aforementioned route from the Northwest Shelf of Australia to Tokyo, Japan as emphasized in Figure 16.

Figure 15: Heat Rate Curve (Input-Output Curve) for a Power Plant



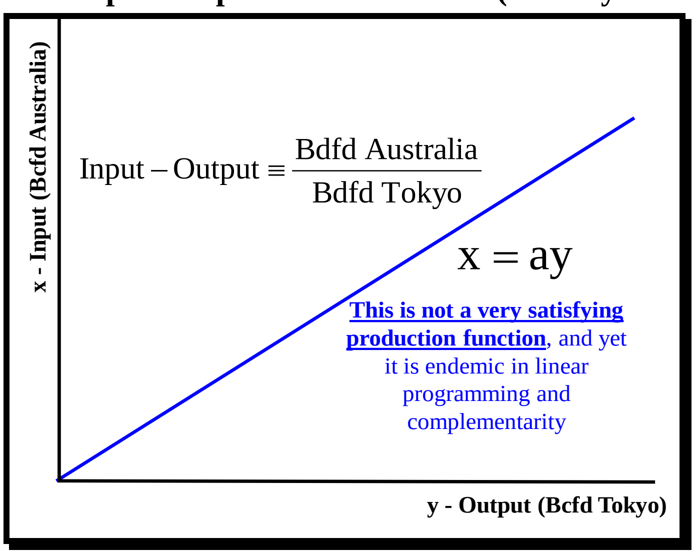
Lest EIA assert "We couldn't do that," keep in mind **EIA is already doing it**. EIA or contractors are estimating losses along every pipeline and LNG tanker route in your entire world model and your entire North American model, existing and prospective, everywhere in the world. You must do so; otherwise, you can never have an international gas model with a requisite transportation and logistics sector. Some of these estimates are very specific to the pipeline or tanker involved because that pipeline or tanker already exists, and the measurements are public or quasi-public. Some of these estimates are generic. A particular pipeline may not be built yet or isn't reported, but you know the diameter and pressure of the line and the spacing and number of compression stations. This isn't rocket science, and it is generally done rather accurately. Fluid flow laws are ubiquitous and predictive, and EIA will do this routinely throughout the world and report the data in its publications and models. It is not acceptable to say: "We don't have the data." You do have the data, and if you don't, EIA people can get it, infer it, or model it.

Figure 16: Input-Output Curve for LNG Delivered from Northwest Shelf Australia to Tokyo, Japan



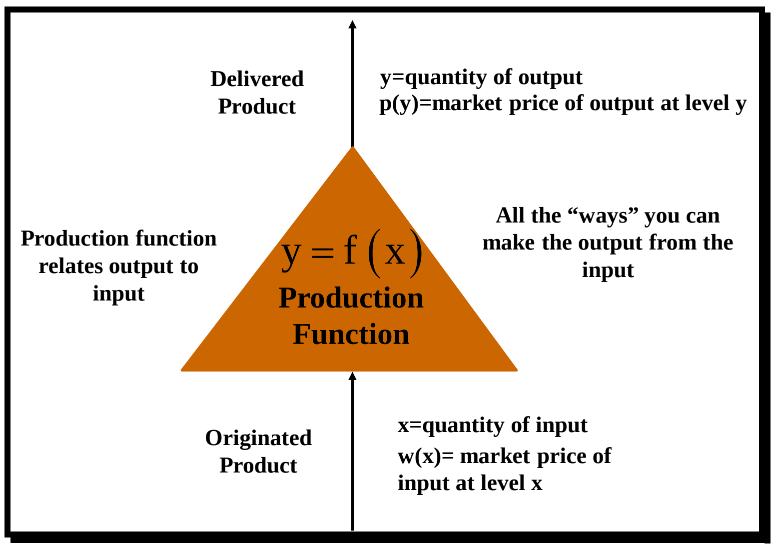
Alas, practitioners of linear programming are compelled to assume a Leontief production function, which means a linear input-output curve. The amount of output you get per unit of input is independent of the magnitude of output. This is the classical Leontief production function. Figure 17 illustrates the Leontief production function, which is a straight line. The slope of the line is the constant, Leontief input-output coefficient a. We want to do much better in our models than Leontief production functions. (Linear programming models are Leontief, and we want to overcome that.)

Figure 17: Leontief Input-Output Curve Is Linear (Used by Linear Programming)



In order to build a detailed and accurate model of transportation, we return to the notion of each transportation activity in the model being embedded in a price system as illustrated in Figure 18. It is important to articulate and emphasize this in some detail so that the reader will understand the salient results about to be developed, results borrowed directly from the graduate economic textbooks like Mas-Colell et. al. op. cit., Jehle et. al. op. cit., and Varian op. cit. The forthcoming development is literally out of the pages of those texts. Notice on the output link that the quantity of output delivered to the terminus of the transportation process is the quantity y. The quantity of input taken from the upstream end of the transportation process is the quantity x. The quantity y and the quantity x are related by the nonlinear production function y=f(x). The price of the output of the transportation system may be a function of the level of output of the transportation system. We see that price designated p(y), specifically depicting its dependence on the level of output y. The situation in which the price of the output p(y) is a function of the level of output y is a case in which the specific pipeline might have and be able to exercise "market power." This is a case of a dominant seller or monopolist. We will return to this theme later.

Figure 18: This General Model of Production Suspended in a Price System



The price of the input to the transportation system may be a function of the level of input of the transportation system. We see that price designated w(x), specifically depicting its dependence on the level of input x. The situation in which the price of the input w(x) is a function of the level of input x is a case in which the specific pipeline might have and be able to exercise "market power" against the people who sell it product for transport. This is the classic case of a monopsony, a dominant buyer. We will return to this theme later in this paper.

With this transportation process suspended in a pricing system with the specific nature here, how does the producer make profit? In particular, what are his decision variables? What are his external variables? The answer to the question is simple:

> yp(y) = revenue from sale of product to market at the downstream end of the transportation process.
>
> xw(x) = cost of input product bought from the market at the upstream end of the transportation process.

ϕy = variable cost of the output y.

The decision variable the transporter has is how much product y to produce and how much input x to consume. Mathematically, and of course in the real world, the profit the producer gets if he chooses to produce the level of output y and consume the levels of input x would be

$$\pi(y, x) = yp(y) - \phi y - xw(x)$$

We use the notation π to represent profit. This is rather common notation in microeconomics.

"This seems trivial! Assuredly this is not sufficiently detailed or robust enough to be a measure of profit in the real world?" Au contraire; it is. In the real world, one obviously embeds more detail in one's representation of profit (e.g., taxation, finance) and more dynamics than this simple static case. However, real world profit is not much more difficult than this simple "revenue minus cost" formulation. This simplified case will develop all the fundamental concepts one needs for an agent-based model and can be easily generalized to deal with real world complexity. One can superimpose taxes, royalties, and other costs to expand this model but the fundamental result remains invariant.

Let us summarize the key market structure concepts:

- **Competitive situation**—p(y) is not a function of y, w(x) is not a function of x, i.e., p(y) = p and w(x) = w.

- **Monopoly situation**—p(y) is an explicit function of y.

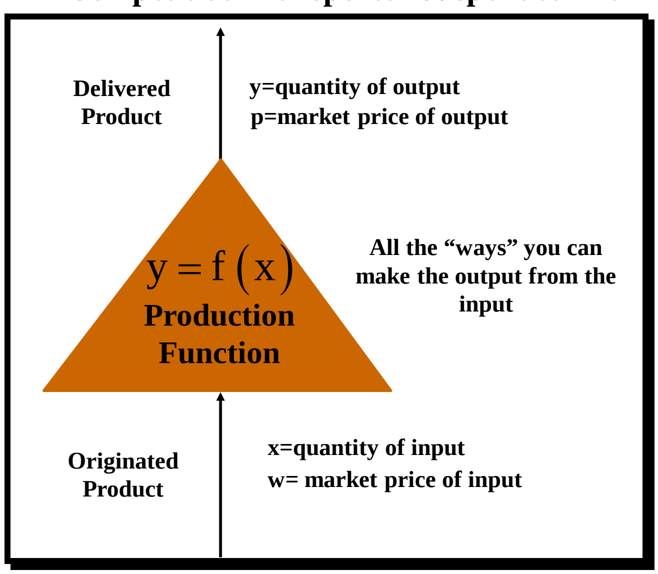
- **Monopsony situation**—w(x) is an explicit function of x.

In order to have a monopoly or monopsony, **there must be sufficient concentration** that a single producer can via his production or consumption decisions **unilaterally** affect price. This is a rare situation in the real world. However, the fundamental microeconomic approach that we put forth here will be able to deal with it if we expect it to be occurring in a particular portion of the world. This gives crystal clear meaning to the concept "market power" rather than some arcane or dilettante meaning. EIA knows this to be fundamentally important. EIA has had to deal with real or prospective market power issues in its modeling for many years (since the days of the "Oil Market Simulation" or OMS model at EIA that was built by Dr. Robert Marshalla, Dale Nesbitt's Decision Focus Incorporated colleague, and later generalized by Dr. Marshalla and Dr. Nesbitt in the form of the DFI World Oil Model for DOE Policy under Jerry Blankenship, Glen Sweetnam, and Steve Minihan.)

We should note that monopoly/monopsony power is difficult to model. Monopoly or oligopoly power is not nearly as trivial as the complementarity monograph by Gabriel op. cit. might have one believe. In fact, the Gabriel op. cit. complementarity book and many other models de facto trivialize it. We will discuss why later in this document; the reason is related to the notion of p(y). The same is true for monopsony as well.

For the main mathematical developments, we will assume a fully competitive solution, with neither monopoly nor monopsony power. The fully competitive solution is the one in which the output price is not a function of the output quantity, and the input price is not a function of the input quantity. In that competitive situation, the transportation process owner must take the price of delivered product at the destination location as given (p), and he or she must take the price of input commodity at the originating location as given (w). We will proceed in detail with the model of a competitive transportation process owner, a transporter who takes input and output prices as given. This is called the "atomistic" assumption in economics. It assumes that the producer does not have the market concentration to affect prices by his unilateral input or output actions and therefore cannot benefit by reducing output or withholding something. In our experience, this is frequently the case throughout North America and the world. It is usually the case that when a producer cuts output, the revenue he receives goes **down**, not up. The revenue (and the price) does not rise faster than the magnitude of his production cutback. (In that situation, the individual producer or transporter has no market power.) This competitive situation is illustrated in Figure 19. (We are going to see that can be equivalent to the complementarity approach outlined in Gabriel op. cit.)

Figure 19: A Competitive Transporter Suspended in a Price System



To reiterate, the competitive transporter profit maximization problem (no monopoly, no monopsony) is one in which the transporter is going to take the price of the product at the destination point as given and the price of the commodity at the originating point as given, and he is going to set his level of purchase at the originating point and level of sale at the destination point so as to maximize his profit. The variables he goes to the market and observes are p and w, the prices of the output and input respectively in his origin and destination markets. The decision variables he has is how much to buy at the origin point (namely x) and how much to sell at the destination point (namely y). That is, the job of the transportation process owner is to select the level of output y and the level of input x that are the solution to the transporter profit maximization problem:

$$
\begin{aligned}
&\text{MAX} \quad py - \phi y - wx \\
&y, x \\
&\text{SUBJECT TO} \quad y = f(x) \\
&y \geq 0 \\
&x \geq 0
\end{aligned}
\qquad \text{Equation 4}
$$

This is a fairly general formulation, more so than it might appear on its face. We have assumed a variable cost per unit of output ϕ, and therefore the total variable cost will be ϕy. Of course, this problem can get more complex as we introduce taxes, royalties, time variation, capital investment, fixed costs, etc., but the fundamentals are truly no more difficult than this. Keep in mind, the production function relates all possible outputs and inputs. The transporter's job is to take prices as given and choose which specific level of output and inputs that maximize his profit. The prices are fixed and constant with respect to the (competitive) transporter's output and input decision.

The crucially important point is this. **The transporter does not have to take account of anyone or anything else**. The only thing that concerns the transporter is this:

- What is the price of the product p at the destination that he serves?
- What is the price of the input w at the origin that he serves?
- What is the nonfuel variable cost ϕ per unit of output?
- What is the production function f(.) that relates quantity of output to quantity of input?

This transporter profit maximization problem concerns only the transporter; no other party in the entire economy is involved at all. This transporter is not marching in goose step lockstep to the beat of a monolithic global welfare function. In the model as well as the real world, there is no way this transporter is forced to march to any drummer other than his own profit measure. This formulation evidences the beauty of microeconomics, a decentralized, agent-based behavioral view of maximizing profits in the face of the pricing system in which the transporter is embedded. There is no motivation whatsoever for any agent to care about anything other than the prices for products and factors he faces.

This transporter profit maximization problem is a mathematical optimization problem. It was solved in complete generality by the mathematicians Kuhn and Tucker in the 1960s. We refer to the solution equations as the "Kuhn-Tucker conditions." Luenberger op. cit. puts forth out the Kuhn-Tucker theory in great detail in an accessible form.

The optimum answer, i.e., the profit-maximizing set of outputs and inputs, is most definitely a function of the product price p and the input price w. That is emphasized by the following graphic:

$$\begin{array}{l} \underset{y,x}{\text{MAX}} \quad (p-\phi)y - wx \\ \text{SUBJECT TO} \quad y = f(x) \\ y \ge 0 \quad x \ge 0 \end{array} \Rightarrow \begin{array}{l} x^*(p,w) \\ y^*(p,w) \end{array}$$

The secret of microeconomics is simple—**<u>solve this problem ANALYTICALLY, not numerically</u>**. That is the mantra—do not solve something numerically that you can solve analytically.

5.1 The Mathematical Solution

This section presents the mathematics of the solution to the profit maximization problem. The first step is to write the problem in Luenberger form:

$$\begin{array}{l} \underset{y,x}{\text{MIN}} \quad -(p-\phi)y + wx \\ \text{SUBJECT TO} \quad y - f(x) = 0 \\ y \ge 0 \\ x \ge 0 \end{array}$$

With the transporter profit maximization problem written in this form, the Lagrangian is:

$$L(y,x,\lambda) = -(p-\phi)y + wx + \lambda[y - f(x)] - \mu_y y - \mu_x x$$

The Kuhn-Tucker conditions are, therefore, the following:

$$\begin{array}{l} \dfrac{\partial L(y,x,\lambda)}{\partial y} = -(p-\phi) + \lambda - \mu_y = 0 \\ \dfrac{\partial L(y,x,\lambda)}{\partial x} = w - \lambda f'(x) - \mu_x = 0 \\ \dfrac{\partial L(y,x,\lambda)}{\partial \lambda} = y - f(x) = 0 \\ \mu_y y = 0 \qquad \mu_y \ge 0 \qquad y \ge 0 \\ \mu_x x = 0 \qquad \mu_x \ge 0 \qquad x \ge 0 \end{array}$$

Substitute the first two equations into the last two equations to eliminate the Lagrange multipliers on the inequality constraints.

$$y - f(x) = 0$$

$$\left[\lambda-(p-\phi)\right]y=0 \quad \lambda-(p-\phi)\geq 0 \quad y\geq 0$$
$$\left[w-\lambda f'(x)\right]x=0 \quad w-\lambda f'(x)\geq 0 \quad x\geq 0$$

If we assume an interior-point solution, x>0 and y>0, we can simplify the foregoing equations to the form:

$$y-f(x)=0$$
$$\lambda=p-\phi$$
$$w=\lambda f'(x)$$

We can substitute out the Lagrange multiplier and write the well-known solution as follows:

$$y=f(x)$$
$$\frac{w}{p-\phi}=f'(x)$$

We are going to need the specific functional form of the production function f(.) if we are to proceed further. Needing a functional form is not at all a problem or disadvantage with the method. The same requirement is true with other methods including monolithic global welfare maximization via complementarity or other methods. You need these production functions sooner or later, and sooner is better.

What is the marginal rate of transformation? Because f(.) has the units of output y, then the function f'(.) has the units y/x. It is how much additional deliverability y you get at the margin for the level of input x that you have. The marginal rate of transformation is the rate at which you convert an input to an output **at the margin**, i.e., the derivative of the production function. The Kuhn-Tucker conditions tell us that the transporter will operate such that the derivative of the production function is exactly equal to the price ratio between input and output. To operate at any other level is **not** to maximize profits. This **always** occurs because if it does not, the transporter is not maximizing his or her profits. Keep squarely in mind, these are the conditions for profit maximization. The solution is the profit maximizing level of input and output from the transportation process. This is not a trivial point, and worthy of belaboring. You do **not** have to implement some numerical method to figure out the profit maximizing solution. The answer lies in the analytical solution, the functional form, itself. The answer is analytical, and you have just solved the analytical problem. **Once you have solved the analytical problem, there is no need for any complementarity or Kuhn-Tucker conditions**. You have already taken full account of them.

5.2 Example from Figure 14

If we used the production function from Figure 14, which was:

$$x=y^{a}\Rightarrow y=x^{\frac{1}{a}}$$

we could write the answers in fully closed form. In this case the production function is:

$$x = y^a \Rightarrow y = x^{\frac{1}{a}} \Rightarrow f(x) = x^{\frac{1}{a}} \Rightarrow f'(x) = \frac{1}{a} x^{\frac{1-a}{a}}$$

Substituting into the foregoing two Kuhn-Tucker conditions, we get the following answer for y > 0 and x > 0:

$$\frac{w}{p-\phi} = \frac{1}{a} x^{*\frac{1-a}{a}} \Rightarrow x^*(p,w) = \left(\frac{p-\phi}{aw}\right)^{\frac{a}{a-1}}$$

$$y^*(p,w) = \left(\frac{p-\phi}{aw}\right)^{\frac{1}{a-1}}$$

In the case where $p-\phi \leq 0$, we know that:

$$x^*(p,w) = y^*(p,w) = 0$$

It is worth writing this solution very carefully for future reference:

$$x^*(p,w) = \begin{cases} \left(\frac{p-\phi}{aw}\right)^{\frac{a}{a-1}} & \text{if } p-\phi > 0 \\ 0 & \text{if } p-\phi \leq 0 \end{cases}$$ **Equation 5**

$$y^*(p,w) = \begin{cases} \left(\frac{p-\phi}{aw}\right)^{\frac{1}{a-1}} & \text{if } p-\phi > 0 \\ 0 & \text{if } p-\phi \leq 0 \end{cases}$$ **Equation 6**

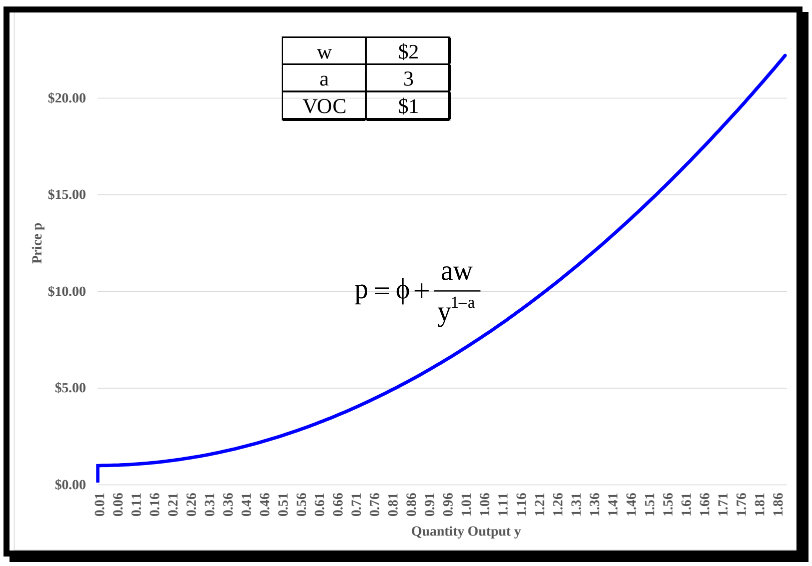
We emphasize by including the arguments the optimum level of output y is a function of p and w, and the optimum level of input y is a function of p and w. The important point to note is that if we specify a production function, we obtain an **analytical** answer to what the profit maximizing level of output y*(p,w) is and what the profit maximizing level of input x*(p,w) is. More importantly, x*(p,w) is going to be interpreted as a demand function for inputs to the transportation process, and y*(p,w) is going to be interpreted as a supply function for outputs from the transportation process. We are going to return to this theme after proving that these types of relationships hold for every production function with which one might wish to characterize the transportation process.

It is illustrative to plot what the output supply curve looks like and what the input demand curve look like for the particular transportation process here. We begin with the output supply curve, and we write the indirect supply function (price on the vertical axis) that people are so used to plotting. We begin by inverting the expression for y*(p,w):

$$y^*(p,w) = \left(\frac{p-\phi}{aw}\right)^{\frac{1}{a-1}} \Rightarrow p = \phi + \frac{aw}{y^{1-a}}$$

This supply function on the output of the transportation node for the particular production function assumed here has the form in Figure 20. This has the rather appealing property that supply is zero for all prices below ϕ. It trends upward when prices are above ϕ with an appealing convex shape. This type of supply function is algorithmically very easy to work with. Supply functions derived from other production function forms are just as easy or easier.

Figure 20: The Supply Curve for Output of the Transportation Process



We write the indirect demand function (price on the vertical axis) that people are so used to plotting. We begin by inverting the expression for x*(p,w)

$$x^*(p,w) = \left(\frac{p-\phi}{aw}\right)^{\frac{a}{a-1}} \Rightarrow w = \frac{p-\phi}{a} x^{\frac{1-a}{a}}$$

This indirect demand function on the input of the transportation node has the form in Figure 21. This has the rather appealing property of being monotonically downward trending with an appealing convex shape. This is a type of demand function with which it is algorithmically very easy to work.

To summarize, the transportation node model from a microeconomic perspective can be summarized mathematically and graphically as in Figure 22.

Figure 21: The Demand Curve for Output of the Transportation Process

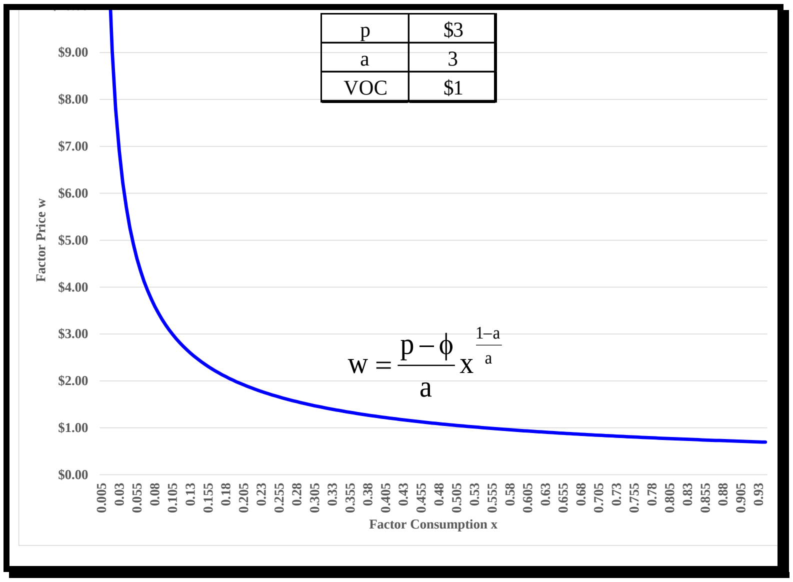
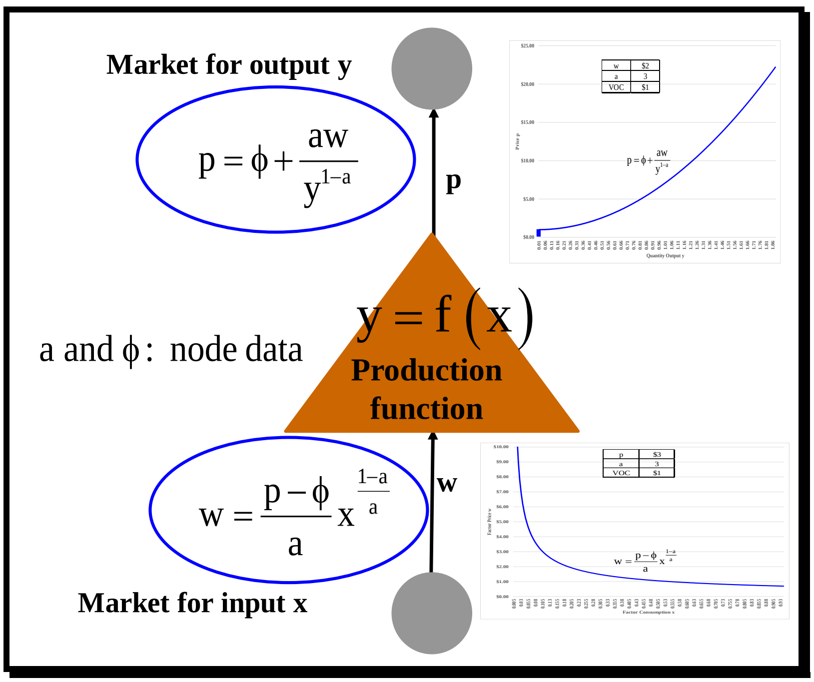


Figure 22: The Microeconomic Model of the Transportation Activity



With this model of transportation, there are no optimization problems to solve, either numerically or analytically. **They've all been solved**! They have all been solved analytically, the best and most desirable way. The only relevant optimization problem to be solved—the transporter profit maximization problem—has already been solved and embedded in the functional forms for the supply and demand curves. The solution is analytic, not numeric. There is no need for a difficult optimization algorithm to be solved every time for this transportation node. There is no need for any complementarity equations—the important equations have already been solved analytically.

The solution is done and complete and expressed as two analytic functions, an analytic supply curve for output and an analytic demand curve for input. In the next section, we are going to give some analytical substance to the supply function on output and the demand function on input, and then we are going to return to the previous production function example.

We should emphasize that we can make these selfsame calculations for any real world production function we might want to apply. The functional equations for the supply curve on the output and the demand curve on the input will be different, but those curves will occur in the same way and have the same generic shapes. There is no implicit limitation whatsoever here.

5.3 Formal Mathematical Proof and Derivation of Supply and Demand Curves

What is the profit that the transporter will receive if he or she sees the price w at the origin and the price p at the destination of his transport process? The answer is simple. He or she will calculate the profit-maximizing input level x*(p,w) and the profit maximizing output level y*(p,w) and will substitute them into his or her measure of profit. Thus, the profit that the transporter achieves facing the prices p and w for output and input respectively is:

$$\pi(p,w) = y^*(p,w)p - x^*(p,w)w$$

Economists call this the "profit function." It is the maximum achievable profit that the transporter will receive if he or she sees the output price p and the input price w. We will develop some rather profound results from this profit function by some mathematical manipulation through a sequence of steps. (The appendix to this section proceeds through an entirely equivalent yet very insightful development from the perspective of the well-known "cost function." The reason we include that appendix is to show precisely and mathematically what the statement "price equals marginal cost" truly means.)

5.3.1 Step 1a: Differentiate the Profit Function with Respect to Input Price w

We use the chain rule to differentiate the profit function with respect to the input price w:

$$\frac{\partial \pi(p,w)}{\partial w} = \frac{\partial y^*(p,w)}{\partial w} p - \left[x^*(p,w) + \frac{\partial x^*(p,w)}{\partial w} w \right]$$

$$= -x^*(p,w) + \left[p\frac{\partial y^*(p,w)}{\partial w} - w\frac{\partial x^*(p,w)}{\partial w} \right]$$

5.3.2 Step 2a: Differentiate the Production Function with Respect to Input Price w

The next step is to substitute the optimum input and output levels into the production function and thereafter differentiate it with respect of input price w. First the substitution:

$$y^*(p,w) - f[x^*(p,w)] = 0$$

and then the differentiation:

$$\frac{\partial y^*(p,w)}{\partial w} - f'[x^*(p,w)]\frac{\partial x^*(p,w)}{\partial w} = 0$$

5.3.3 Step 3a: The Kuhn-Tucker Conditions Must Hold for the Optimum Solution

The Kuhn-Tucker conditions must hold for the optimum solution. We substitute the optimum solution into the Kuhn-Tucker conditions developed from the profit maximization problem:

$$w - pf'\left[x^*(p,w)\right] = 0$$

$$y^*(p,w) = f\left[x^*(p,w)\right]$$

We arrange the first equation to solve for f’[]:

$$f'\left[x^*(p,w)\right] = \frac{w}{p}$$

5.3.4 Step 4a: Substitute the First Kuhn-Tucker Equation from Step 3a into Step 2a Result

We substitute the derivative from step 3a into the expression containing the derivative in step 2a. This substitution will yield the result:

$$\frac{\partial y^*(p,w)}{\partial w} - f'[x^*(p,w)]\frac{\partial x^*(p,w)}{\partial w} = 0$$
$$\Rightarrow \frac{\partial y^*(p,w)}{\partial w} - \frac{w}{p}\frac{\partial x^*(p,w)}{\partial w} = 0$$

If we multiply this equation by the price p, we obtain:

$$p\frac{\partial y^*(p,w)}{\partial w} - w\frac{\partial x^*(p,w)}{\partial w} = 0$$

5.3.5 Step 5a: Step 4a Eliminates the Bracketed Term in Step 1a

Based on Step 4a, we see that the bracketed term in Step 1a is zero, i.e.:

$$\frac{\partial \pi(p,w)}{\partial w} = -x^*(p,w) + \left[p\frac{\partial y^*(p,w)}{\partial w} - w\frac{\partial x^*(p,w)}{\partial w}\right]$$

and therefore:

$$\frac{\partial \pi(p,w)}{\partial w} = -x^*(p,w)$$

The negative of the derivative of the profit function with respect to the price of the input is the demand for the input. This is a rather profound definition of the demand curve for the input at the point of origin of the transportation process. This result, used extensively throughout microeconomics, is a result of the Envelope Theorem of optimization. In fact, we have actually re-proved the envelope theorem in the foregoing development. We now turn to the supply curve for output.

5.3.6 Step 1b: Differentiate the Profit Function with Respect to Output Price p

Let us use the chain rule and differentiate the profit function with respect to the output price p:

$$\frac{\partial \pi(p,w)}{\partial p} = \left[y^*(p,w) + \frac{\partial y^*(p,w)}{\partial p} p \right] - \frac{\partial x^*(p,w)}{\partial p} w$$
$$= y^*(p,w) + \left[p \frac{\partial y^*(p,w)}{\partial p} - w \frac{\partial x^*(p,w)}{\partial p} \right]$$

5.3.7 Step 2b: Differentiate the Production Function with Respect to Output Price p

The next step is to substitute the optimum input and output levels into the production function and thereafter differentiate it with respect to output price p. First the substitution:

$$y^*(p,w) - f[x^*(p,w)] = 0$$

and then the differentiation:

$$\frac{\partial y^*(p,w)}{\partial p} - f'[x^*(p,w)] \frac{\partial x^*(p,w)}{\partial p} = 0$$

5.3.8 Step 3b: The Kuhn-Tucker Conditions Must Hold for the Optimum Solution

The Kuhn-Tucker conditions must hold for the optimum solution. We substitute the optimum solution into the Kuhn-Tucker conditions developed from the profit maximization problem:

$$w - pf'\left[x^*(p,w)\right] = 0$$

$$y^*(p,w) = f\left[x^*(p,w)\right]$$

We arrange the first equation to solve for f'[]:

$$f'\left[x^*(p,w)\right] = \frac{w}{p}$$

5.3.9 Step 4b: Substitute the First Kuhn-Tucker Equation from Step 3b into Step 2b Result

Substitute the derivative from step 3a into the expression containing the derivative in step 2a. This substitution will yield the result:

$$\frac{\partial y^*(p,w)}{\partial p} - f'[x^*(p,w)]\frac{\partial x^*(p,w)}{\partial p} = 0$$
$$\Rightarrow \frac{\partial y^*(p,w)}{\partial p} - \frac{w}{p}\frac{\partial x^*(p,w)}{\partial p} = 0$$

If we multiply this equation by the price p, we obtain:

$$p\frac{\partial y^*(p,w)}{\partial p} - w\frac{\partial x^*(p,w)}{\partial p} = 0$$

5.3.10 Step 5b: Step 4b Eliminates the Bracketed Term in Step 1b

Based on Step 4a, we see that the bracketed term in Step 1a is zero, i.e.:

$$\frac{\partial \pi(p,w)}{\partial p} = y^*(p,w) + \left[p\frac{\partial y^*(p,w)}{\partial p} - w\frac{\partial x^*(p,w)}{\partial p}\right]$$

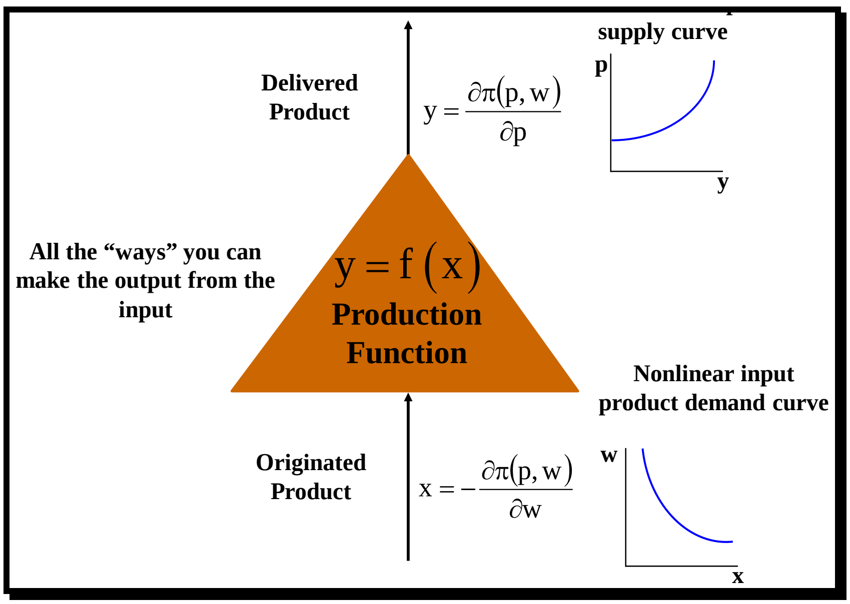
and therefore:

$$\frac{\partial \pi(p,w)}{\partial p} = y^*(p,w)$$

The derivative of the profit function with respect to the price of the output is the supply of output. This is a rather profound definition of the supply curve for the output at the point of delivery from the transportation process. This result, used extensively throughout microeconomics, is a result of the Envelope Theorem of optimization. We have re-proved the envelope theorem again here.

We now have the demand curve for input and the supply curve for output of this transportation process, and we have developed them specifically from the profit function. We have appended them to the transportation process diagram from Figure 13 to emphasize what we have just proven. We have proven that the transporter profit maximization problem derives and places an analytic supply curve on the delivered product provided by the transportation process, and it simultaneously derives and places an analytic demand curve on the origin location product taken in by the transportation process. There is a supply curve on output, and there is a demand curve on input. All that is required to get them is the profit function, and all that is required to get the profit function is the production function. If you specify an analytical form for the production function, the supply curve on output and the demand curve on input are analytical functions, and they are absolutely, totally, and unequivocally consistent with profit maximization on the part of the transporter. There are absolutely no complementarity or other subsequent calculations required. The optimization is finished, complete, and embedded within the functional forms. There is no need for complementarity, Kuhn-Tucker, linear programming, or any such equations or complexities. They have all been analytically solved.

Figure 23: A Supply Curve for Delivered Product at the Destination and the Demand Curve for Product at the Origin



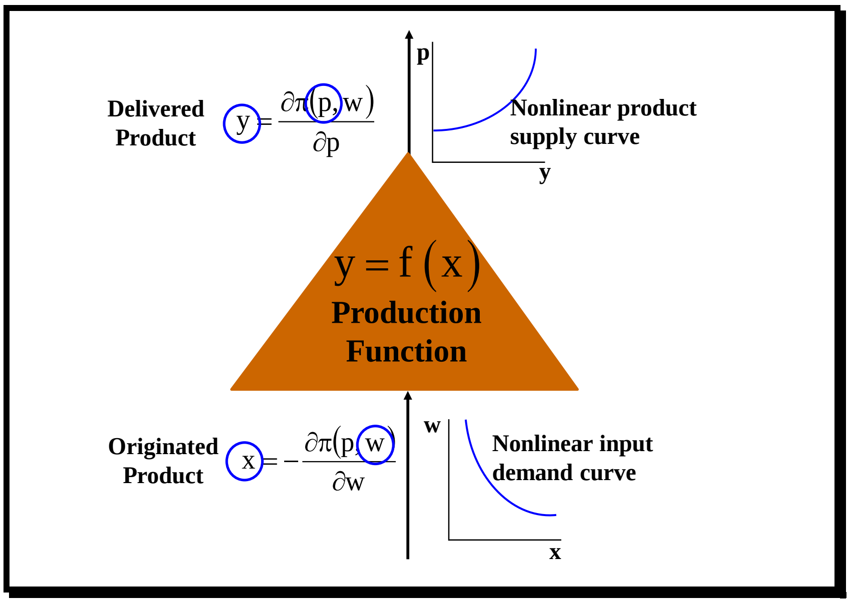
These curves are derived in total from the profit function, and they are completely **analytical**. There is no optimization problem that has to be solved here; it has already been solved. It is the profit maximization problem and the profit maximization problem alone that has derived these supply and demand curves in the first place! There is no need for any Kuhn-Tucker conditions or complementarity equations; they have been solved analytically. This is a profound insight, for it allows us to build a model that represents profit-seeking behavior on the part of transporters without ever having to resort to a numerical or other optimization procedure at any transportation node in one's model. The optimization is already done by applying the Kuhn-Tucker conditions to the analytical production function and solving the optimization problem **analytically**.

It is very useful to show diagrammatically in Figure 24 why these are, in fact, true and correct supply functions on output links and demand functions on input links. It is important to notice that the supply curve from the top expresses y as a function of p, but lurking there is also the dependence on input price w. This means that the supply curve for output is a function of input price, and also the demand curve for input is a function of output price. This is not only true, it is expected. It is this connection that is going to unite and connect an entire network together in a very simple, natural, and solvable way.

The crucial insight is this. Profit maximization and profit maximization **alone** puts:

- a **supply curve** on every output link.
- a **demand curve** on every input link.

Figure 24: Why These Are Truly Supply Curves and Demand Curves Respectively



I

5.4 The Profit Function for the Example in Figure 14

We calculate the profit function for the example in Section 4.2. If we substitute the optimum values of x*(p,w) and y*(p,w) into the expression for profit (p-ϕ)y-wx, we can write:

$$wx^*(p,w) = \begin{cases} w\left(\frac{p-\phi}{aw}\right)^{\frac{a}{a-1}} & \text{if } p-\phi > 0 \\ 0 & \text{if } p-\phi \le 0 \end{cases}$$

$$(p-\phi)y^*(p,w) = \begin{cases} (p-\phi)\left(\frac{p-\phi}{aw}\right)^{\frac{1}{a-1}} & \text{if } p-\phi > 0 \\ 0 & \text{if } p-\phi \le 0 \end{cases}$$

and therefore:

$$\pi(p,w)=(p-\phi)y^*(p,w)-wx^*(p,w)=\begin{cases} w(a-1)\left(\dfrac{p-\phi}{aw}\right)^{\frac{a}{a-1}} & \text{if } p-\phi>0 \\ 0 & \text{if } p-\phi\le 0 \end{cases}$$

If one differentiates this profit function, one obtains the supply and demand functions exactly as proven and calculated previously.

$$\frac{\partial\pi(p,w)}{\partial p}=\begin{cases} w(a-1)\dfrac{a}{a-1}\left(\dfrac{p-\phi}{aw}\right)^{\frac{a}{a-1}-1}\dfrac{1}{aw}=\left(\dfrac{p-\phi}{aw}\right)^{\frac{1}{a-1}} & \text{if } p-\phi>0 \\ 0 & \text{if } p-\phi\le 0 \end{cases}$$

To write the derivative with respect to w, first isolate the w term in the factor demand equation:

$$\pi(p,w)=\begin{cases} (a-1)(p-\phi)^{\frac{a}{a-1}}a^{-\frac{a}{a-1}}w^{-\frac{a}{a-1}+1} & \text{if } p-\phi>0 \\ 0 & \text{if } p-\phi\le 0 \end{cases}$$

Thereafter:

$$\frac{\partial\pi(p,w)}{\partial w}=\begin{cases} (a-1)(p-\phi)^{\frac{a}{a-1}}a^{-\frac{a}{a-1}}\left(-\dfrac{a}{a-1}+1\right)w^{-\frac{a}{a-1}} & \text{if } p-\phi>0 \\ 0 & \text{if } p-\phi\le 0 \end{cases}$$

$$=\begin{cases} -\left(\dfrac{p-\phi}{aw}\right)^{\frac{a}{a-1}} & \text{if } p-\phi>0 \\ 0 & \text{if } p-\phi\le 0 \end{cases}$$

This shows that all you need in each transportation node is a fully optimized, analytical profit function. Creating profit functions from production functions that represent the performance of transportation processes is easy, a veritable homework problem in microeconomics with solutions widely known, understood, and published. This simple analytic solution saves an extreme amount of solution time and complexity as we shall see. It will obviate the need for solving or Kuhn-Tucker conditions altogether. The solution has been found analytically.

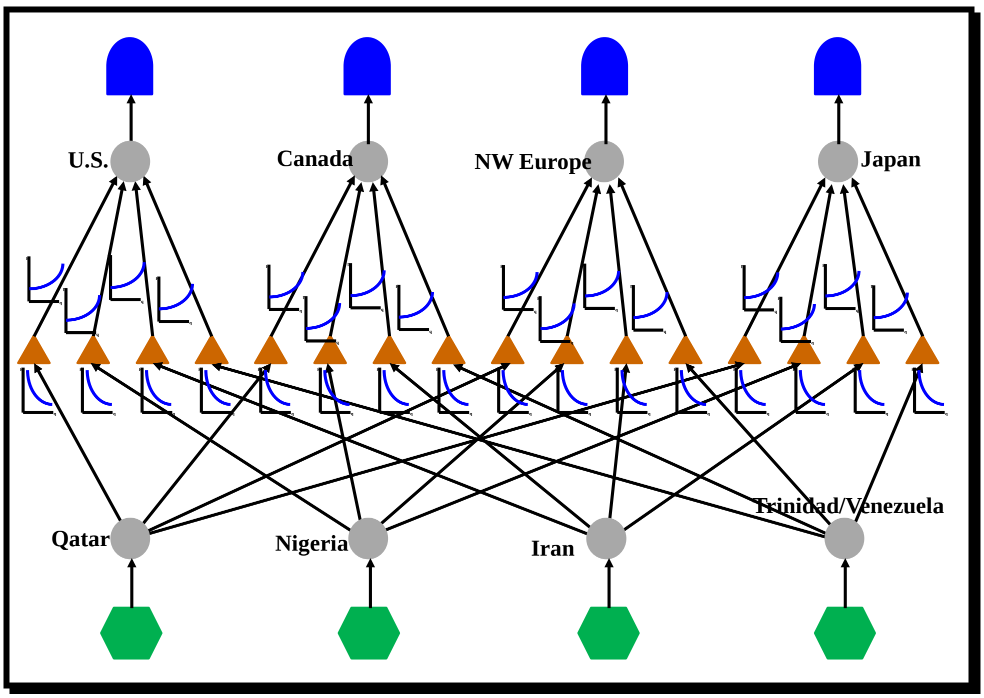
5.5 This Structure Suggests a Nodes and Links Network Formalism

This finding that each transportation process has a supply curve at is point of delivery and has a demand curve at its point of origin suggests a very powerful nodes and links network formalism. Returning to the example transportation problem in Figure 1, we have hereby determined that there is a supply curve on the output of every transportation process, and we have calculated it

analytically. No numerical algorithms. No linear programs. No complementarity. No unsolved Kuhn-Tucker conditions of any type. No iterative schemes. No algorithms. No difficult iterative methods to solve optimization problems fraught with size and difficulties.

Similarly, there is a derived, analytical, functional form for the supply curve pointing out the top of every transportation node in the network that works every time it is calculated. Similarly, there is a demand curve on the input to every transportation process, and we have calculated it **analytically**. Yet again, there are no numerics at all required for that. No linear programming. No complementarity. No unsolved Kuhn-Tucker conditions of any type. No painful iterative methods to solve optimization problems fraught with risks and difficulties. There is a derived, analytical, functional form for every demand curve pointing out the bottom of every transportation node in the network. The analytical situation is that in Figure 25. Look how progress toward the solution has already been made just by analytically solving the transporter profit maximization Kuhn-Tucker conditions. Thus far, there are absolutely no numerical solutions or calculations that have to be made.

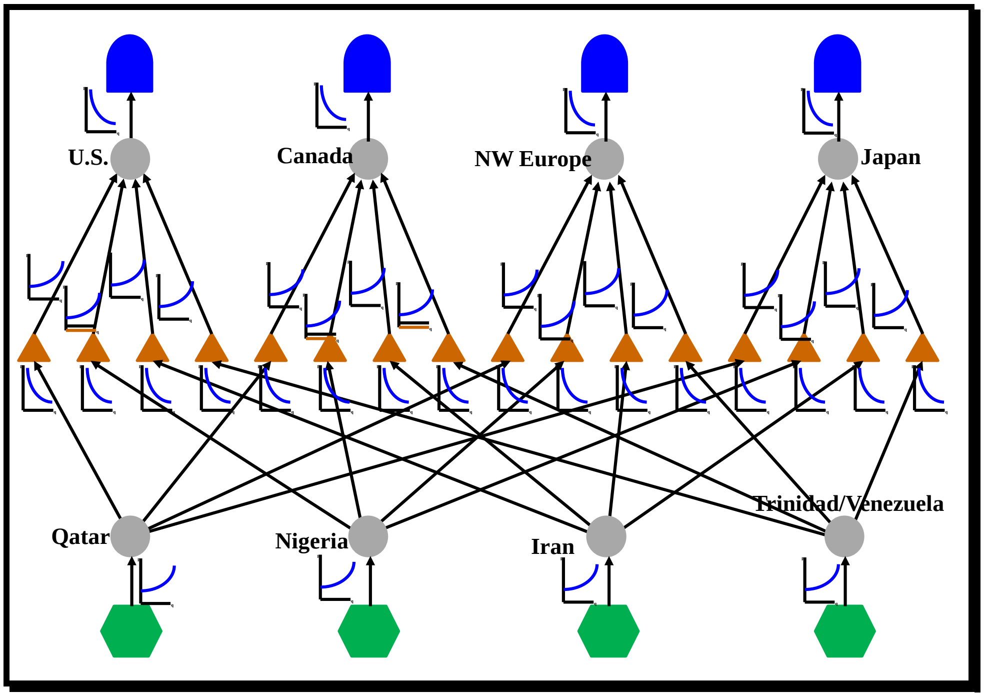
Figure 25: All of the Transportation Activities with Supply Curves on Output and Demand Curves on Input



Without offering mathematical proofs, let us assume that there is a demand curve on the input link to each demand node in Figure 25. That is hardly a “big stretch.” That indeed is the function of a demand curve—to put demand onto the input link into that demand node. Let us also assume that there is a supply curve on the output link to each supply curve node in Figure 25. That is hardly a “big stretch” either. That indeed is the function of a supply curve—to put supply onto the output link from that supply node. We have appended these supply curves and demand curves to the transportation-only representation in Figure 25, creating Figure 26. Notice in Figure 26:

- There is a **demand curve** on the input link to every economic node.
- There is a **supply curve** on the output link from every economic node.
- These supply curves and demand curves are derived from profit maximization.
- Their very existence **implies** profit maximization. Their equation structure already reflects profit maximization.

Figure 26: Supply Curves and Demand Curves Appended to Supply and Demand Links



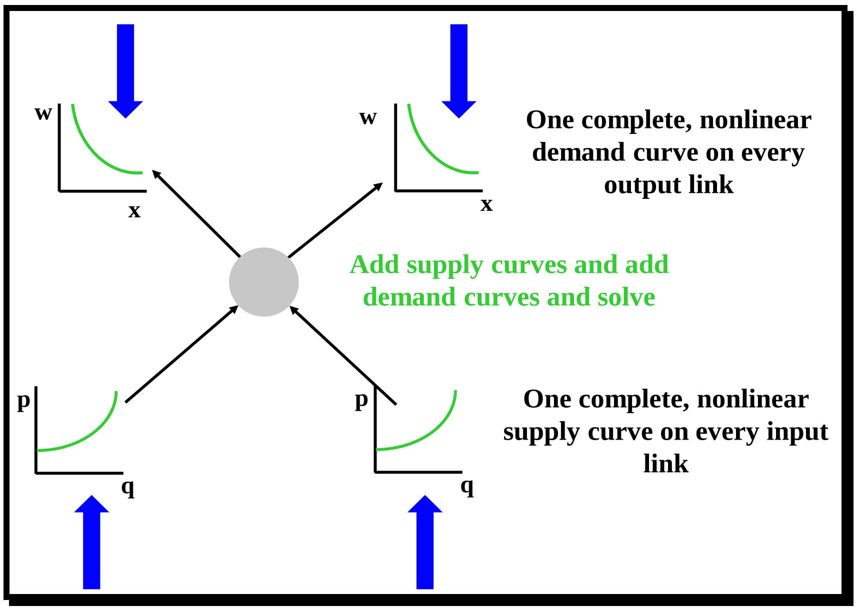
Notice further that every economic node in the model (supply and transportation) has an output link that terminates in a market hub for each of its outputs. That is, every economic node delivers to a market hub. Furthermore, every economic node in the model (demand and transportation) has an input link that originates in a market hub for its sources. That is, every economic node originates from a market hub for each of its inputs. This situation for the economic nodes—supply, transportation, and demand are now very, very simple:

- Every resource supply node has a supply curve for its output.
- Every demand node has a demand curve for its input.
- Every transportation node has a supply curve for its output and a demand curve for its input.
- All these supply and demand curves are fully and completely analytical.

The key to solving this transportation problem is to switch our attention to the hubs in the network. What does every hub see? It sees, as emphasized in Figure 27, a demand curve coming downward into it from every economic or demand node immediately above, and it sees a supply curve coming upward into it from every economic or supply node immediately below. This construct absolutely "buys you the farm" analytically and allows one to solve the problem easily, completely, and generally.

If every hub in the network sees a supply curve on every input link coming into it, then horizontal addition of supply curves into that hub means the hub sees a "supply stack" comprised of the aggregate supply curve over all the input links. In short, the sum of all inputs to the hub is a supply stack obtained by horizontal addition as in Figure 28. Horizontal addition simply means at each given level of price, one adds the quantities on the direct supply curves for each of the direct supply curves coming into the hub. How simple is this? How direct is this? How non-numerical is this? It is very simple, very direct, and very non-numerical. There is no need for complementarity or other overkill here. The problem has been rendered much simpler, extremely simple in fact, by virtue of the analytical solutions that have occurred.

Figure 27: Every Hub Marries Aggregate Nodal Supply and Aggregate Nodal Demand



The aggregate supply curve coming into the bottom of the market hub is the lower right envelope of the individual supply curves in Figure 28, as represented in the thick line at the lower right in Figure 29.

It is insightful to redact the individual, competing supply curves in Figure 29 to draw the single aggregate supply curve in Figure 30. That is where the "action" is going to be algorithmically, the aggregate supply curve coming into the hub.

Figure 28: Horizontal Addition of Supply Curves Creates Aggregate Nodal Supply Stack

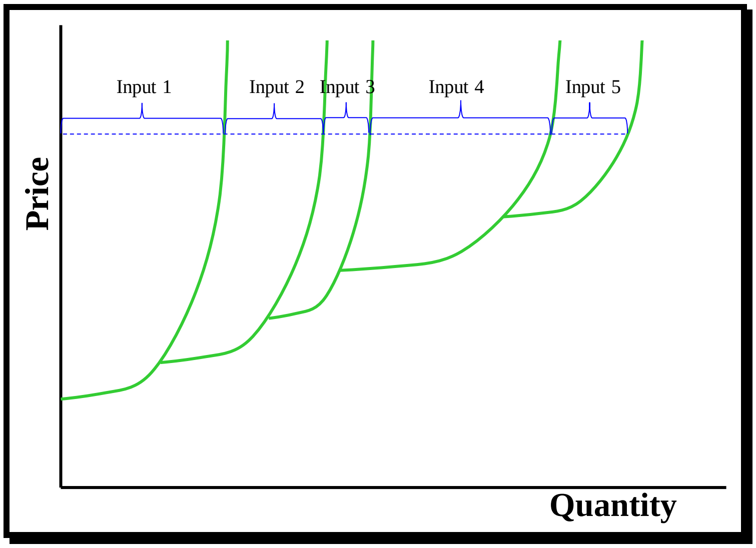
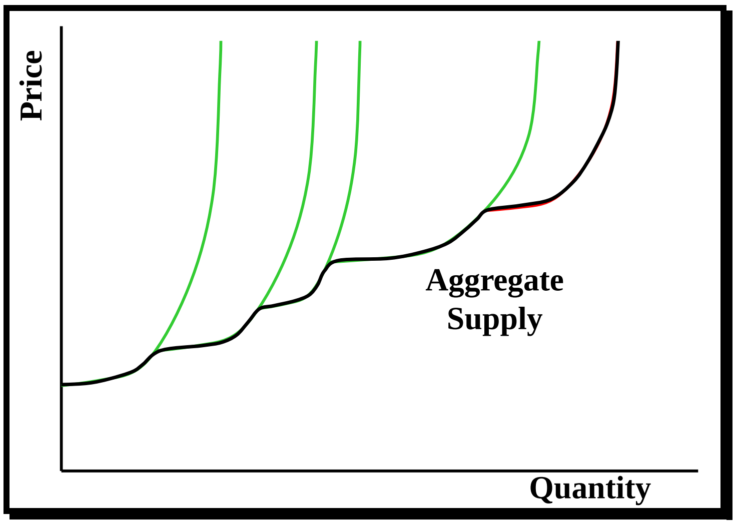


Figure 29: Horizontal Addition of Supply Curves Creates an Aggregate Nodal Supply Stack



Turning to the demand side, if every hub in the network sees a demand curve on every output link going out of it, then horizontal addition at that hub means the hub sees a "demand stack" comprised of the aggregate on all output links. In short, the sum of all outputs from each hub is a demand stack obtained by horizontal addition as in Figure 31. Horizontal addition simply means at each given level of price, one adds the quantities on the direct demand curves for each of the direct demand curves leaving the hub. As before, how simple is this? How direct is this? How non-numerical is this? It is very simple, very direct, and very non-numerical indeed.

Figure 30: The Aggregate Supply Curve Coming into the Hub

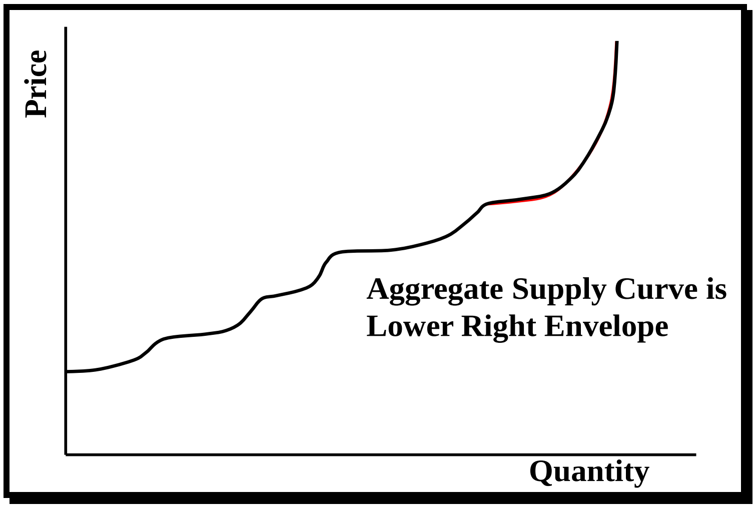
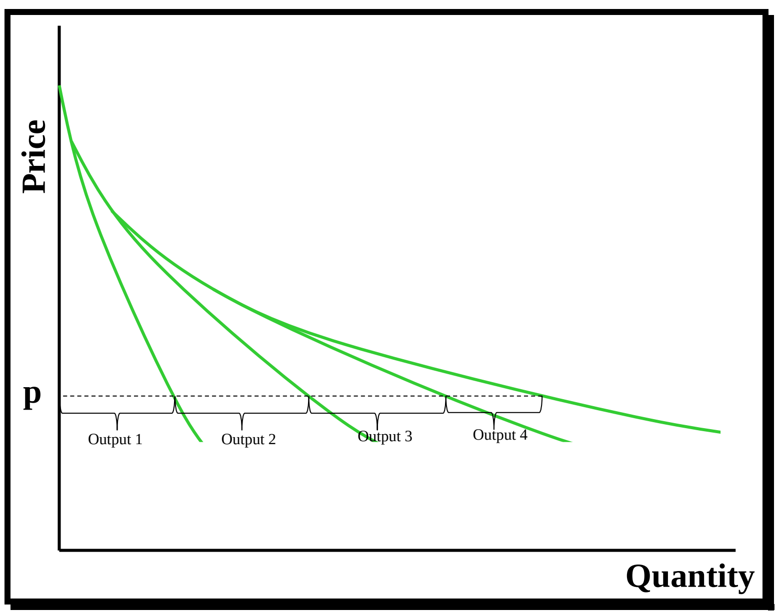


Figure 31: Horizontal Addition of Demand Curves Creates Aggregate Nodal Demand Stack



The horizontal sum of the demand curves is an aggregate demand stack coming out of the hub node. We indicate the horizontal sum of the demand curves in Figure 32.

It is illustrative to redact the individual, competing demand curves in Figure 32 to draw the single aggregate demand curve. That is where the "action" is going to be algorithmically, the aggregate demand curve coming out of the hub. The single aggregate demand curve at the hub appears in Figure 33.

Figure 32: Horizontal Addition of Demand Curves Creates Aggregate Nodal Demand Stack

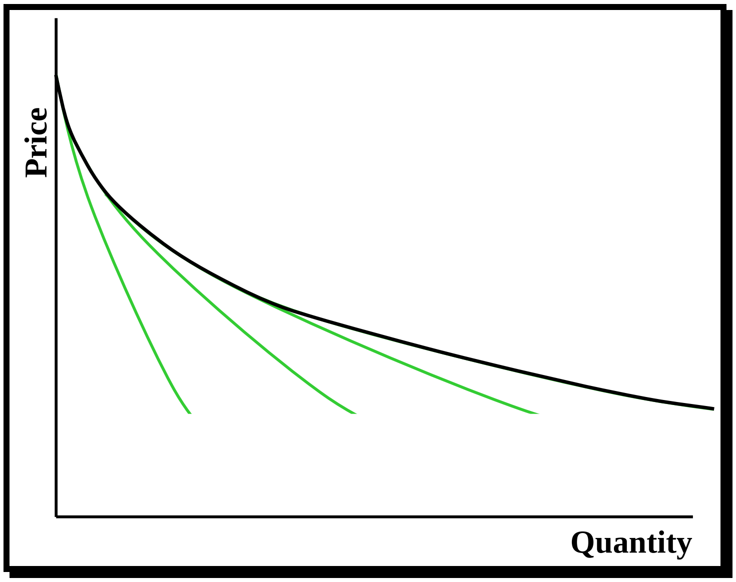
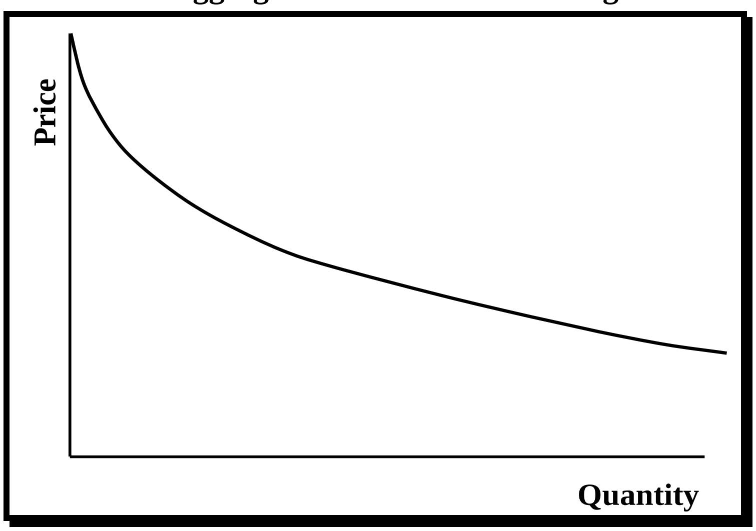
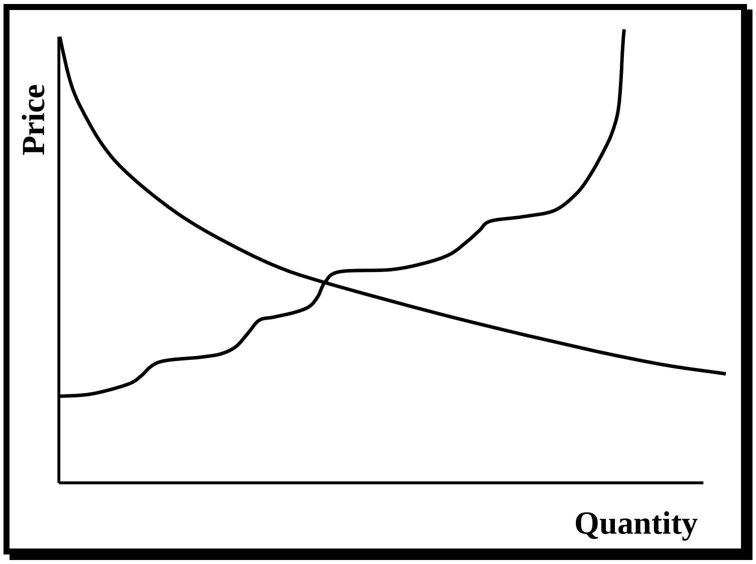


Figure 33: The Aggregate Demand Curve Going Out of the Hub



Conveniently, every hub in the network sees an aggregate supply curve coming in on the one or more input links, and it sees an aggregate demand curve going out on the one or more output links. Therefore, every hub in the network sees the supply-demand curve pair shown in Figure 34. This is all possible by virtue of the fact that we have **analytically solved the Kuhn-Tucker conditions** for every economic activity in the entire network, and those analytically solved Kuhn-Tucker conditions result in a supply curve on every output of every economic node, and they result in a demand curve on every input of every economic node. This is a literally magical property, for it allows and admits of very easy, straightforward, and constructive solution of the problem at every hub in the model. We are already getting a glimpse how much easier this is than a complex, full rank complementarity solution that takes no account of analytical solutions to any part of the problem.

Figure 34: Every Hub in the Network Has a Supply-Demand Curve Pair

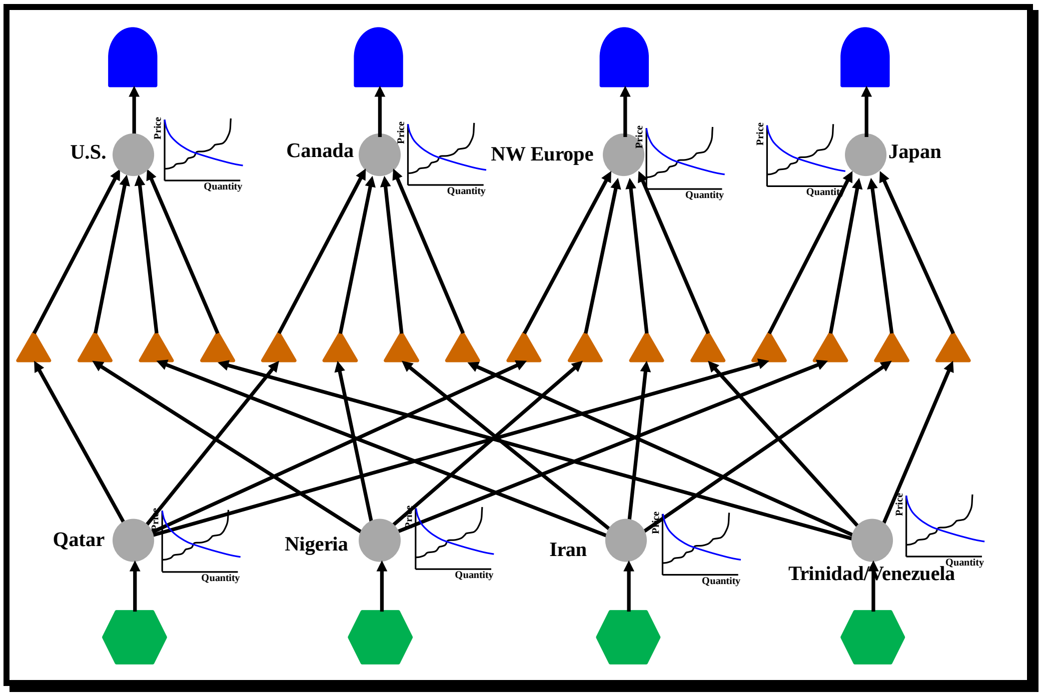


What does this mean in terms of the original network in Figure 26? The answer is emphasized in Figure 35, which depicts an aggregate supply curve and an aggregate demand curve at every hub in the model. How hard is this problem to solve algorithmically? Not hard at all. In fact, the solution is extremely simple. There is certainly no reason to search for a complex, arcane solution algorithm or “program up” a bunch of equations to “deliver them to a solver.” Complementarity and linear programming are out of the game, far too large and complicated relative to this. It is crystal clear why the solution has been reduced from the Cartesian product of hubs plus activities times hubs plus activities (the original complementarity problem size) to simply hubs (the new problem size). The activities have all been solved **analytically**. That activity dimension is solved out! That is the key—analytical solutions allow you to solve out the economic activity dimension. You are analytically rather than numerically solving each activity.

A simple and direct algorithm to solve this simple problem in Figure 35 would be the following:

- Initialization: Make a guess at the price and quantity in the U.S.
- Initialization: Make a guess at the price and quantity in Canada.
- Initialization: Make a guess at the price and quantity in NW Europe.
- Initialization: Make a guess at the price and quantity in Japan.
- Initialization: Make a guess at the price and quantity in Qatar.
- Initialization: Make a guess at the price and quantity in Nigeria.
- Initialization: Make a guess at the price and quantity in Iran.
- Initialization: Make a guess at the price and quantity in Trinidad and Venezuela.

Figure 35: An Aggregate Supply Curve and an Aggregate Demand Curve at Each Hub in the Model



You can initiate a guess in any way you want, including randomization. Then you continue as follows:

- Iteration 1: Solve for the crossing point price and quantity in the U.S.
- Iteration 1: Solve for the crossing point price and quantity in Canada.
- Iteration 1: Solve for the crossing point price and quantity in NW Europe.
- Iteration 1: Solve for the crossing point price and quantity in Japan.
- Iteration 1: Solve for the crossing point price and quantity in Qatar.
- Iteration 1: Solve for the crossing point price and quantity in Nigeria.
- Iteration 1: Solve for the crossing point price and quantity in Iran.
- Iteration 1: Solve for the crossing point price and quantity in Trinidad and Venezuela.

- Iteration 2: Solve for the crossing point price and quantity in the U.S.
- Iteration 2: Solve for the crossing point price and quantity in Canada.
- Iteration 2: Solve for the crossing point price and quantity in NW Europe.
- Iteration 2: Solve for the crossing point price and quantity in Japan.
- Iteration 2: Solve for the crossing point price and quantity in Qatar.
- Iteration 2: Solve for the crossing point price and quantity in Nigeria.
- Iteration 2: Solve for the crossing point price and quantity in Iran.
- Iteration 2: Solve for the crossing point price and quantity in Trinidad and Venezuela.

If all crossing point prices and quantities are the same in Iteration k as they were in Iteration k-1, stop. You have found the network equilibrium. If not, do another iteration. Continue until all crossing point prices and quantities are the same on successive iterations.[29]

Why do you have to iterate successively? It is because as we have seen, each and every transportation node has the property that the supply curve for output is a function of not only the output price but also the input price. Simultaneously, the demand curve for input is a function of not only the price of the input but also the price of the output. In an extremely important way, the transportation processes connect the entire system together. That should be no surprise; **that is exactly what transportation in fact does in the real world!** This requires that we successively solve the simple supply-demand problems at the nodes in our model until all are completely consistent and unchanging from iteration to iteration. This will fully and totally interconnect the entire network and achieve a full network equilibrium, which is precisely what we want.

5.6 Solution Methods Are Simple, Reliable, and Well Known

How would one solve a supply-demand curve pair like that in Figure 34? It is not rocket science! It is, in fact, a very well understood problem and quite simple and direct algorithmically. It is after all a one-dimensional problem, not a vector problem. Methods for solving one-dimensional supply-demand curve pair crossings are well known and anything but controversial. The methods are right out of the economics and mathematics books, and they are not linear programming, nonlinear programming, or complementarity. We reiterate that complementarity is not even mentioned in any graduate microeconomics book; we have not found a single graduate textbook in microeconomics that mentions complementarity, global welfare maximization, or any such operations research technique as a potential solution method. (And that is not because complementarity is new or "high tech." It has been around since the days of Kuhn and Tucker and Samuelson in 1952.) Such discussions are concentrated in the operations research literature, and they have not been relevant to microeconomic considerations. Microeconomics has soundly rejected even the slightest hint of a monolithic global welfare function. Microeconomics wants to be totally agent-based, and microeconomics wants prices to consider prices, not shadow prices. Microeconomics wants zero probability of pollution or distortion of the answer.

How would you solve a simple supply-demand curve pair? There are several proven methods to choose from:

- Newton's method
- Secant method
- Fixed point method
- Cobweb method
- Walras excess demand method

---

[29] This is not the algorithm used in ArrowHead (nor in MarketBuilder). The ArrowHead algorithm is more sophisticated and more parallelizable than this. However, this algorithm works and illustrates the power of the network microeconomic approach. Parenthetically, this is only one of many ways to solve the eight interconnected supply-demand curve pairs. There are many other methods, including full rank Newton's methods. A full rank Newton's method here would be **substantially** smaller than the ones for complementarity.

5.6.1 One-dimensional Newton's Method

With this method, we would write the supply function as q=S(p) and we would write the demand function as q=D(p). The excess demand would be z(p)=D(p)-S(p). We want z(p) to be 0, so we want to solve the equation:

$$D(p) - S(p) \triangleq z(p) = 0$$

Newton's method iterates in the equation (with or without relaxed adjustment:

$$p_{k+1} = p_k - \frac{z(p_k)}{z'(p_k)}$$

until successive prices are equal. It is as if a fictitious auctioneer was yelling out the price p_k on the kth iteration and then on the k+1 iteration adjusting it according to the second term.

5.6.2 One-dimensional Secant Method

The secant method is quite similar to the Newton method. The secant method linearizes the excess demand curve z(p) and uses a linear approximation in place of the derivative. All else is equal to Newton's method.

5.6.3 One-dimensional Fixed Point Methods

Fixed point methods begin with a price p. They then calculate the supply at that price q=S(p). They then calculate the inverse demand at that quantity p=D(q). They then combine those two mappings to assert:

$$p = D\left[S(p)\right] = f(p)$$

They then iterate in the equation (with or without relaxation) until the price does not change from iteration to iteration:

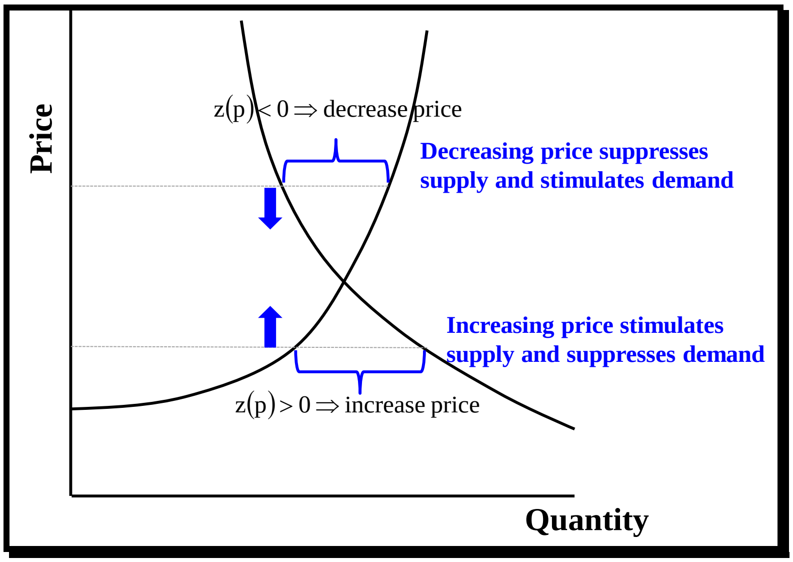
$$p_{k+1} = f(p_k)$$

For a single supply-demand curve pair, the fixed point algorithm is no more than an economic cobweb algorithm. This algorithm, and the acceleration techniques that guarantee it to work, are well known and very effective and efficient.

5.6.4 Auction Approach

We could use an auction method in which a price is posited, the excess demand is calculated, and the price is adjusted based on the excess demand. This is quite akin to the Newton's algorithm above, which adjusts the price p_k by an adjustment factor every iteration. The diagram of an auction algorithm is very simple, as in Figure 36.

Figure 36: Auction Algorithm



There is a major benefit here because the solution methods are well known, simple, and known to work perfectly and reliably for real world supply and demand curves (particularly monotonically increasing supply and monotonically decreasing demand). These are highly desirable methods for another crucial reason—they are entirely local. They do not require a giant matrix that spans the entire system (**no Jacobean matrix** or equivalent), which linear programming and complementarity do. There are only 8 supply-demand curve problems, one for each hub. This is a far smaller system than the simplified 32 by 32 equations and unknowns system that characterize complementarity even for this simple transportation network (much less the 64 equations and unknowns before the simplification in Section 4). Unlike the ponderous and sparse complementarity and linear programming formulations, unidimensional supply-demand algorithms are completely local and extremely easy to solve. That is the beauty of the network equilibrium method. Local, extremely easy solution algorithms repeated until the global solution is achieved. No Hessians. No Jacobians. No complexity or programming difficulty. No multidimensional Newton's methods. No multidimensional algorithms. No "conditioning" problems. No singularity problems. No bundling of equations to deliver them to a "solver."

Furthermore, and this is crucially important, there are **only eight** equations and unknowns to be solved. There is no matrix with rows equal to the number of hubs and columns equal to the number of activities. The columns are all gone, **replaced by analytic solutions to transporter profit maximization equations**. That entire dimension of complexity is eradicated from the problem, having been solved analytically in generality, before the fact, by solving the transporter profit-maximization problem. Any time one can substitute an analytic solution for a numerical solution, that is precisely the thing to do because one achieves such a high degree of problem solving gain and compaction. The difficulty with complementarity and linear programming, aside from the fact that they maximize monolithic global welfare, is that we have not seen them ever attempt to consider analytic solutions. If there is any hint in the wind of an optimization problem, complementarity defaults immediately to solving it numerically as part of a huge concentrated algorithm, cramming it into a monolithic solution method or "solver" (e.g., GAMS, AMPL) using labor intensive and error prone GAMS-like input-output formats. Complementarity does not strive to solve it analytically, and that is a veritable coup de grace for the method.

5.7 Algorithmic Parallelization

The other crucially important point is algorithmic parallelization. Recall the first iteration of the algorithm above:

- Iteration 1: Solve for the crossing point price and quantity in the U.S.
- Iteration 1: Solve for the crossing point price and quantity in Canada.
- Iteration 1: Solve for the crossing point price and quantity in NW Europe.
- Iteration 1: Solve for the crossing point price and quantity in Japan.
- Iteration 1: Solve for the crossing point price and quantity in Qatar.
- Iteration 1: Solve for the crossing point price and quantity in Nigeria.
- Iteration 1: Solve for the crossing point price and quantity in Iran.
- Iteration 1: Solve for the crossing point price and quantity in Trinidad and Venezuela.

Why couldn't one put each of these solutions onto an individual microprocessor (or "thread") and run the processors simultaneously in parallel? There are eight problems and they would run in 1/8 of the time! Just assign each of the hubs to a thread or a processor and let them run individually in parallel. You could let them run synchronously or asynchronously; it does not matter. That is precisely what you would want to do, and this is what we do. We get the solutions in 1/8 of the time. This is algorithmic parallelization. You systematically assign each node to its own processor and let the processors all run in parallel. There is nothing serendipitous about such parallelization; it is systematic and algorithmic. You achieve it by assignment and scheduling of hubs to processors. Unlike other approaches, you do not allow the serendipity of the compiler to hopefully figure it out on the fly.

To conceive the power of algorithmic parallelization such as this, suppose that you had 16,000 hubs in your transportation problem, a huge, world size, inclusive transportation matrix, and you had 8 processors.[30] You would assign 2,000 hubs to each processor and run those 2,000 bundles of hubs simultaneously in parallel. The situation with multiple nodes (rather than individual nodes) assigned to each processor is no different. You would get the solution in 1/8 of the time, excluding any inter-node overhead. You do not want to "hope and pray" for parallelization based on "serendipity" of the processor, seeking to advance incremental tasks that it might on its own volition identify to advance. (That is the best that monolithic, serial codes such as GAMS or AMPL complementarity or optimization solvers can hope for—serendipitous parallelization scheduled by the compiler according to its own internal algorithms, which are invisible to you as the operator.

[30] Our present ArrowHead Global Oil and Refined Products Model has approximately this level of detail.

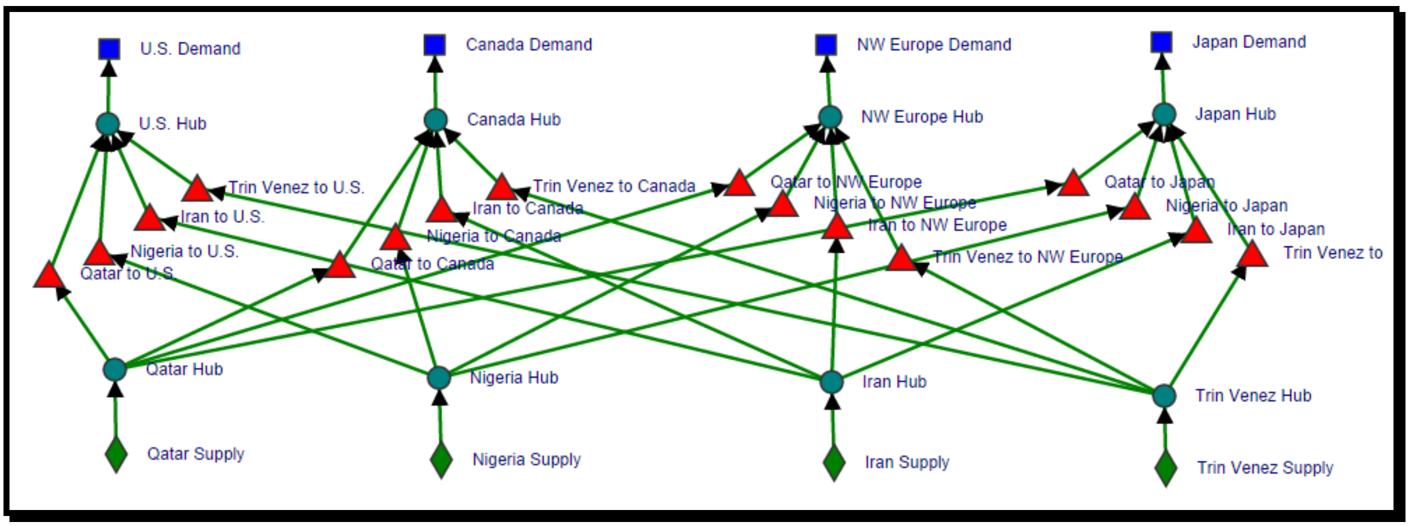
That has not proven to give much acceleration on serial algorithms, which move from point-to-point based on information from the previous point.) On the contrary, we want to assign nodes to processors and systematically and algorithmically control the parallelization. This ensures that one actually achieves the maximum parallelization! We do not rely on compiler serendipity; we rely on cold, hard, rational, systematic assignment and execution algorithms to every thread or processor on the computer.

5.8 How Do You Build a Network Model?

Building a network model of the type shown here is easy and quick—in our experience at least two orders of magnitude quicker and more accurate than complementarity or linear programming.

We have included the representation of the network diagram as represented in the ArrowHead software system in Figure 37. Once this network is entered into the ArrowHead software system, this network diagram **is** the model. Readers now know exactly why. All you have to do is enter the data for each of the production functions for each of the transportation nodes (an easy task), enter the demand curves in the demand nodes, and enter the supply curves in the supply nodes. Then you click "Solve" and the model solves in the way we have just seen in very short order using the highly transparent algorithm articulated herein.

Figure 37: The Example Network As Represented in ArrowHead



5.9 Summary

With network microeconomics, all one solves is simple, single variable supply demand curve pairs. Such solutions are trivial. For the example in this report, this is a mere 8 interconnected, simultaneous equations and unknowns. With complementarity, you solve for a minimum of 32 simultaneous equations and unknowns using a difficult, full rank, fully interconnected, monolithic, Jacobean-based multivariate Newton algorithm. Which is faster? Which is easier? Which is more transparent? Which gives interim and final results that are more meaningful? Which uses analytic solutions in place of numerical solutions to dramatically reduce the size and complexity of the algorithm, and in fact to eliminate the need for complementarity types of solutions altogether? To ask the question is to answer it.

In conclusion, we encourage the reader to try a number of different production functions and proceed through the developments here, solve the problem, and see what shapes of output supply functions and input demand functions one obtains. It is highly elucidating. We should emphasize that we use dynamic, intertemporal production functions with particularly convenient and descriptive shapes within ArrowHead. Those functions and shapes are exposed to our users in complete detail, as is the mathematics to derive them. However, it is fundamentally no more difficult or challenging than what has been proven in this section with the simple production function.

5.10 Appendix 5.1—Cost Function Approach

We return to the transporter profit maximization problem in Equation 4 above, rewritten here for convenience:

$$\begin{aligned} &\underset{y,x}{\text{MAX}} \quad py - \phi y - wx \\ &\text{SUBJECT TO} \quad y = f(x) \\ &y \geq 0 \\ &x \geq 0 \end{aligned} \qquad \text{Equation 7}$$

Economists noted a long time ago that this problem can be divided into two hierarchical parts and written as follows:

$$\begin{array}{c} \text{MAX} \\ \text{SUB. TO} \quad y \geq 0 \end{array} \quad py - \left\{ \begin{array}{l} \text{MIN} \quad \phi y + wx \\ \text{SUB. TO} \quad y = f(x) \\ x \geq 0 \end{array} \right\} \qquad \text{Equation 8}$$

They have noted that the profit maximizing strategy involves first the transporter to minimize his or her cost of producing each given level of output y given the input factor price w and thereafter setting the level of output so as to maximize profit. The profit maximizing strategy requires acquiring factors for any given level of output at the lowest possible cost. Economists have then gone further to define the “cost function” or the “total cost function” to be the solution to the cost minimization problem on the right:

$$\begin{aligned} TC(y,w) \triangleq &\underset{x}{\text{MIN}} \quad \phi y + wx \\ &\text{SUBJECT TO} \quad y = f(x) \\ &x \geq 0 \end{aligned} \qquad \text{Equation 9}$$

Thereafter, the full profit maximization problem can be rewritten:

$$\begin{aligned} &\text{MAX} \quad py - TC(y,w) \\ &y \\ &\text{SUBJECT TO} \quad y \geq 0 \end{aligned} \qquad \text{Equation 10}$$

This section will focus initially on the cost function in Equation 9, and then we will solve the full profit-maximization problem in Equation 10. The cost-minimization problem in Equation 9 is already in Luenberger form, so we can write its Lagrangian as follows:

$$L(y,w) = \phi y + wx + \lambda\left[y - f(x)\right] - \mu x$$

The Kuhn-Tucker conditions for the cost minimization problem in Equation 9 are:

$$\begin{aligned} &\frac{\partial L(y,w)}{\partial x} = w - \lambda f'(x) - \mu = 0 \\ &\frac{\partial L(y,w)}{\partial \lambda} = y - f(x) = 0 \\ &\mu x = 0 \quad \mu \geq 0 \quad x \geq 0 \end{aligned} \qquad \text{Equation 11}$$

If we examine Equation 9, we notice that the solution, i.e., the optimizing value of the input factor x, is a function of the input price w and the output level y. Thus, we write the answer to the optimum for Equation 9:

$$x^*(y,w)$$

We substitute into the cost of producing the output y at the factor price w to write the total cost the transporter will bear to provide the output y at the input price w:

$$TC(y,w) \triangleq \phi y + wx^*(y,w)$$

This is the minimum achievable cost that the transporter will bear if he or she sees the output quantity y and the input price w. We will be assuming that the assumed level of output y >0 so that the level of input x > 0. The zero output case is simple thereafter.

5.10.1 Step 1a: Differentiate the Cost Function with Respect to Input Price w

We use the chain rule to differentiate the profit function with respect to the input price w:

$$\frac{\partial TC(y,w)}{\partial w} \triangleq x^*(y,w) + w\frac{\partial x^*(y,w)}{\partial w}$$

5.10.2 Step 2a: Differentiate the Production Function with Respect to Input Price w

The next step is to substitute the optimum input and output levels into the production function and thereafter differentiate it with respect of input price w. First the substitution:

$$y - f[x^*(y,w)] = 0$$

and then the differentiation:

$$\frac{\partial y}{\partial w} - \frac{\partial f[x^*(y,w)]}{\partial w} = 0 = -f'\left[x^*(y,w)\right]\frac{\partial x^*(y,w)}{\partial w}$$

5.10.3 Step 3a: The Kuhn-Tucker Conditions Must Hold for the Optimum Solution

The Kuhn-Tucker conditions must hold for the optimum solution. Because y>0, we know that x>0 and thus μ=0. We substitute the optimum solution into the Kuhn-Tucker conditions developed from the profit maximization problem:

$$w - \lambda f'\left[x^*(y,w)\right] = 0$$

$$y^*(p,w) = f\left[x^*(p,w)\right]$$

We arrange the first equation to solve for f'(.):

$$f'\left[x^*(y,w)\right] = \frac{w}{\lambda}$$

5.10.4 Step 4a: Substitute the First Kuhn-Tucker Equation from Step 3a into Step 2a Result

We substitute the derivative from step 3a into the expression containing the derivative in step 2a. This substitution will yield the result:

$$0 = -f'\left[x^*(y,w)\right]\frac{\partial x^*(y,w)}{\partial w}$$

$$\Rightarrow \frac{w}{\lambda}\frac{\partial x^*(y,w)}{\partial w} = 0$$

If we multiply this equation by the Lagrange multiplier λ, we obtain:

$$w\frac{\partial x^*(y,w)}{\partial w}=0$$

5.10.5 Step 5a: Step 4a Eliminates the Bracketed Term in Step 1a

Based on Step 4a, we see that the right-hand term in Step 1a is zero, i.e.:

$$\frac{\partial TC(y,w)}{\partial w}\triangleq x^*(y,w)$$

This has a delightful parallelism with the profit function, as economists have noticed over the years. The derivative of the cost function with respect to the price of the input is the demand for the input. This is a rather profound definition of the demand curve for the input at the point of origin of the transportation process. This result, used extensively throughout microeconomics, is a result of the Envelope Theorem. We have again re-proved the envelope theorem in the foregoing development. We now turn to the supply curve for output.

5.10.6 Step 1b: Differentiate the Profit Function with Respect to Output Price p

We are halfway home, but we must now turn to the full profit maximization problem in Equation 10. We can write it in Luenberger form:

$$\begin{aligned}&\text{MIN} \quad -py+TC(y,w)\\&y\\&\text{SUBJECT TO} \quad -y\le 0\end{aligned} \qquad \text{Equation 12}$$

The Lagrangian for this problem is:

$$L(y)=-py+TC(y,w)-\mu y$$

Let us write the Kuhn-Tucker conditions:

$$\frac{\partial L(y)}{\partial y}=-p+\frac{\partial TC(y,w)}{\partial y}-\mu=0$$

$$\mu y=0 \quad \mu\ge 0 \quad y\ge 0$$

In the situation in which y>0, we have:

$$p=\frac{\partial TC(y,w)}{\partial y}$$

The derivative of the cost function with respect to the quantity of the output is the price of output. This is a rather profound definition of the supply curve for the output at the point of delivery from the transportation process. This result, used extensively throughout microeconomics, is a result of the Envelope Theorem of optimization. We have re-proved the envelope theorem again here. This is the hackneyed old result that **price equals marginal cost**. One can clearly see from this development that price equals marginal cost only happens when transporters are maximizing profit. That is how it was derived. The minute you use a supply curve like that for output, you have already assumed profit maximizing behavior on the part of transporters and other producers. You do not need any Kuhn-Tucker conditions or complementarity after that.

In the situation in which $\mu>0$, we have $y=0$ and $x=0$ and:

$$p < \frac{\partial TC(0, w)}{\partial y}$$

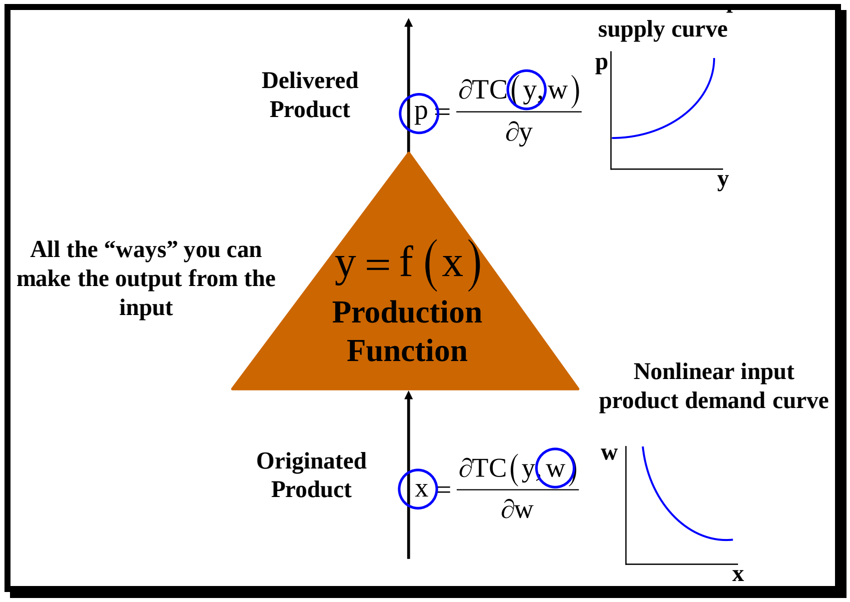
We now have the demand curve for input and the supply curve for output of this transportation process. We have appended them to the transportation process diagram from Figure 13 to emphasize what we have just proven. We have proven that the transporter profit maximization problem derives and places an analytic supply curve on the delivered product provided by the transportation process and it simultaneously derives and places an analytic demand curve on the origin location product taken in by the transportation process. There is a supply curve on output, and there is a demand curve on input. All that is required to get them is the cost function, and all that is required to get the cost function is the production function. If you specify an analytical form for the production function, the supply curve on output and the demand curve on input are analytical functions, and they are absolutely, totally, unequivocally consistent with profit maximization and cost minimization on the part of the transporter. There are absolutely no complementarity or other subsequent calculations required. The optimization is done, complete, and embedded within the functional forms. There is no need for complementarity, Kuhn-Tucker, or any such equations. They have all been analytically solved.

These curves are derived in total from the cost function, and they are completely **analytical**. There is no optimization or complementarity problem that has to be solved here; it has already been solved. It is the cost minimization problem and the profit maximization problem alone that have derived these supply and demand curves in the first place! This is a profound insight, for it allows us to build a model that represents profit-seeking behavior on the part of transporters without ever having to resort to a numerical, or other optimization procedure, at any transportation node in one's model. The optimization is already done by applying the Kuhn-Tucker conditions to the analytical production function and solving the optimization problem analytically.

The crucial insight recurs. Profit maximization and profit maximization **alone** puts:

- a **supply curve** on every output link.
- a **demand curve** on every input link.

Figure 38: A Supply Curve for Delivered Product at the Destination and the Demand Curve for Product at the Origin



5.10.7 The Cost Function for the Example

Returning to the example that derives from Figure 14, we substitute that production function into the cost minimization problem in Equation 12:

$$
\begin{aligned}
TC(y,w) \triangleq\ & \underset{x}{\text{MIN}} \quad \phi y + wx \\
& \text{SUBJECT TO} \quad x = y^a \\
& x \geq 0
\end{aligned}
$$

The solution to this is trivial by substitution:

$$TC(y,w) = \phi y + wy^a$$

It is easy to verify the supply and demand curves by differentiating this cost function per our results here.

$$p = \frac{\partial TC(y,w)}{\partial y} = \phi + way^{a-1}$$

$$x = \frac{\partial TC(y,w)}{\partial w} = y^a$$

If one solves the first equation for y as a function of p and then substitutes it into the second equation, one obtains precisely the same equations as we developed with the profit function. In ArrowHead, we generally rely on the cost function for our demand and supply curves. We actually implement a multi-temporal extension with capacity additions of this rather simple cost function idea. Nonetheless, the concepts here are completely general. The only differences lie in the details.

5.11 Appendix 5.2: Can One Build a Transportation Node with Limited (Constrained) Output?

EIA wants to consider models in which the capacity or the output from a particular transportation node or group of nodes might be limited, i.e., constrained? It is pretty darn easy to do so by adopting a production function that considers capacity (installed capacity) as one of the inputs.

How can one build a model with installed capacities in it as a constraint, i.e., one cannot produce more than some exogenously specified level of capacity c. If we considered a supply function on output of the following form

$$q = c\left[1 - e^{-\alpha(p-\phi_0)}\right]$$

in which

- c is the (constrained) capacity in place (e.g., 2 Bcfd)
- ϕ_0 is the variable cost of the process at zero output (e.g., \$3/Mcf)
- α is a "curvature" parameter (e.g., 5)
- a is an input-output coefficient, expressing the units of input that are required to produce one unit of output.

If we plot this supply curve, we obtain the diagram in Figure 39. We notice that this supply curve is zero at price \$3. By the time price gets to \$3.20, the output is 1.26. By the time the price gets to \$3.40, the output is 1.73. By the time the price gets to \$4, the output is 1.98, virtually 100 percent of capacity. The output is asymptotic to the maximum installed capacity of two units. Just a simple supply curve such as this serves to constrain the output from this transportation process (or other economic process) to two units or lower. This type of production function allows EIA to survey installed capacities and shapes of short term supply curves and input their functional form and their data as easily as this curve was calculated.

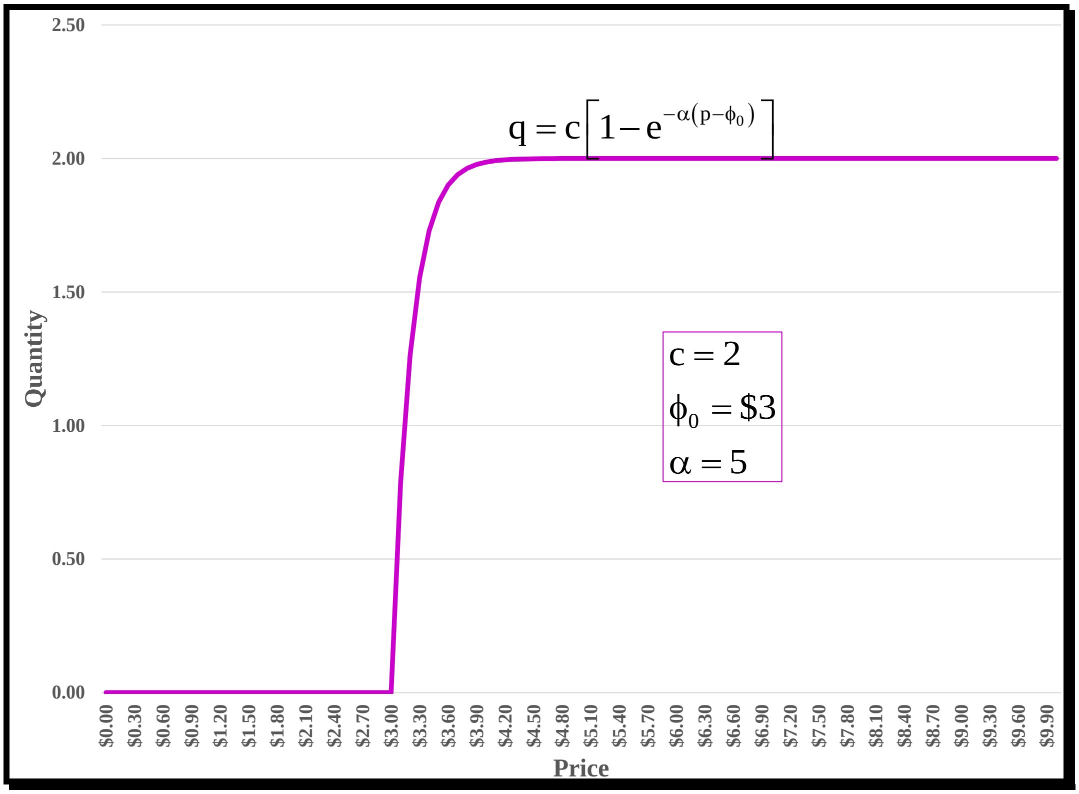
The foregoing equation is the expression for the direct supply curve. The indirect supply curve, calculated by merely inverting the previous function, is

$$p = \phi_0 - \frac{1}{\alpha}\ln\left(\frac{c-q}{c}\right)$$

If we wanted to use this indirect supply curve, what would the Total Cost function have to be? We easily calculate it by integration

$$TC(q,w)=\int_0^q\left[\phi_0-\frac{1}{\alpha}\ln\left(\frac{c-x}{c}\right)\right]dx+wqa=(\phi_0+wa)q+\frac{1}{\alpha}\left[q+(c-q)\ln\left(\frac{c-q}{c}\right)\right]$$

Figure 39: Capacity Constrained Supply Curve



If we calculate the first derivative of this total cost function with respect to q, we obtain the price equals marginal cost relationship we assumed from the outset. The Total Cost function and the Marginal Cost function return the supply curve that we wanted.

$$\frac{\partial TC(q,w)}{\partial q}=(\phi_0+wa)-\frac{1}{\alpha}\ln\frac{c-q}{c}$$

If we insert this Total Cost function into our capacity constrained supply node, Voila!, we obtain a constrained transportation node. This allows EIA to study constrained outputs on transportation process (or in fact to any process to which we apply this logic. In ArrowHead, we have an upper bound function with precisely this form and philosophy, and it allows us to impose upper bounds on individual nodes and collections of nodes at will throughout the entire network model. All we have to do is specify a maximum magnitude of capacity in place (a “not to exceed” capacity) for one or more nodes, and those one or more nodes will thereby be constrained to a maximum total capacity. This works fantastically well for phenomena such as aggregate limitations on LNG tanker fleet.

We see by differentiating the Total Cost function with respect to the input price w, we obtain the demand curve for input

$$\frac{\partial TC(q,w)}{\partial w} = aq = ac\left[1 - e^{-\alpha(p-\phi_0)}\right]$$

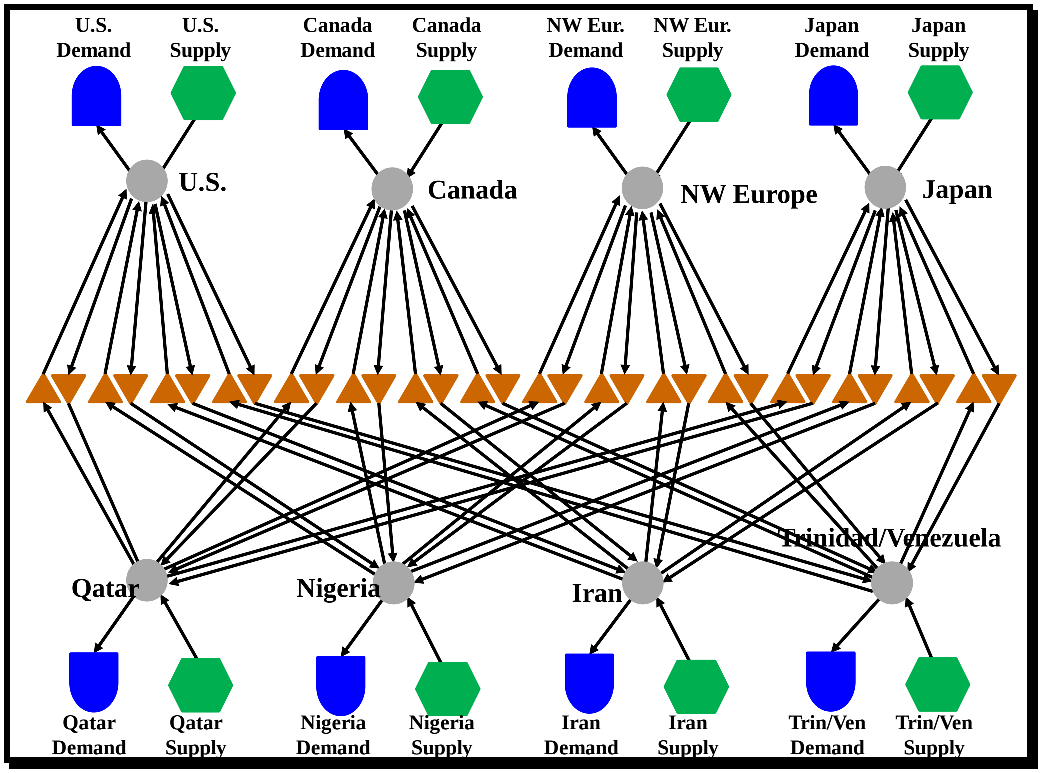
Interestingly, for this particular capacity constrained case, the demand curve for input x is inelastic with respect to the price w of factors. The demand is the same no matter what the factor price w. However, it is a strong function of the output price p. That is precisely the way we put together this illustrative example, and it makes perfect sense. (This is not the only way a capacity-constrained production function can occur. It is a manifestation of the specific production function considered for this illustration, and it works very well for specifying constraints owing from capacity bounds.) It is the constrained output of this process that drives the quantity of both output and input.

It is highly instructive to note that both the supply curve on output and the demand curve on input are fully specified for this capacity constrained production function, and by simply inserting this constrained process into the network, EIA will be able to constrain groups of nodes or single nodes in terms of output by simply varying the installed capacity c. This type of logic will constrain all or part of the transportation grid and allow EIA to represent and study congestion.

6 WORD COMES DOWN FROM ON HIGH: "YOUR TRANSPORTATION MODEL ISNT GOOD ENOUGH; YOU HAVE TO EXPAND IT."

Now comes the inevitable difficulty. Management shows up at your door. *"Things have changed. Massive quantities of oil and gas have been found in formerly importing countries such as the United States and Canada. The 'from-to' nature of transportation has been found to be incorrect. The United States needs both a supply and demand curve because the United States could just as well export LNG as import LNG. Canada needs both a supply and demand curve because Canada could just as well export LNG as import LNG. The same can be said of every other region of the world. Management needs to know about LNG export. Management needs an expanded model that considers import and export everywhere. We need it right now. Congress itself is asking about the level of exports that could occur and what sorts of policies should be considered."* The expanded model must have the structure in Figure 40.

Figure 40: Word Comes Down—We Need to Consider Exports <u>and</u> Imports from Every Country



What do we hear from complementarity or linear programming modelers? *"I need several months to write all the equations, organize them, deliver them to the 'solver,' deliver all the data to the 'solver,' debug the equations and data, make sure they are getting in right, reorganize all the old and new equations, etc."* To change from the original network in Figure 1 to the expanded, bidirectional network in Figure 40 takes weeks or months. There is no easy, straightforward, structured, organized way to do that with complementarity. Complementarians have to write more equations. Complementarians have to amend equations already written. Complementarians have to **<u>program</u>**. Complementarians have to use line by line, opaque, command lines to enter equations and data. They have to insert and read data into line by line data command statements. The complementarity solution algorithm has to grow approximately 50 percent in size. Complementarians have to “bundle” equations and “send them to solvers.”

Network microeconomic people have to do none of these difficult things. They just drag and drop new network elements (transportation processes in this example) and solve the more detailed network problem immediately. They have to write **no** programs. There are still only 8 hubs, so their problem has not grown in size (because all the transportation nodes, old and new, have already been solved analytically). The number of unknowns (the 8 prices at the 8 hubs) has not changed. There are more supply and demand equations entering and exiting each hub, but the difficulty of the problem has not changed materially from the original problem.

What to network microeconomic modelers say? *“I’ll have it done by noon today.”* They don’t have to write any computer code at all. They just drag 12 new transportation nodes into the diagram and drop them in the appropriate location. They drag 24 new links connecting these nodes to the hubs and drop them onto the diagram. They drag in four new supply nodes and drop them on the diagram. They link those four new supply nodes to their local hubs by dragging in links. They drag in four new demand nodes and drop them on the diagram. They link those four new demand nodes to their appropriate local hubs. Total elapsed time is perhaps 15 minutes. Then they type the transportation input-output function data into the 12 new transportation nodes using Excel. They type the supply curve data into the four new supply nodes using Excel. They type the demand curve data into the four new demand nodes using Excel. Total elapsed time is perhaps an hour. They hit the “Run” key, and the same eight supply and demand curve iterative calculation initiates, all in parallel on individual threads. A short time later they have the answer. They phone their manager at 11:15 a.m. that day and say: *“J.C., let’s go to lunch and chat about the new results. We have a day or two to review them before cementing them for management.”*

That is the main reason for using microeconomic network equilibrium. **EIA doesn’t want to write computer code to build a model, to modify a model (redacting structure or adding structure), or to run a model.** EIA wants to write and vet all computer code well in advance and embed it in supply, transportation, and demand nodes with their analytical supply and demand curves before they ever use them. They would want to drag and drop them into a network and use a simple, transparent, effective algorithm to solve them. That in a nutshell is one of the key motivations for microeconomic network modeling of transportation (and other economic activities). You never write one single line of code to build or modify a model, nor do you send anything to a “solver.” You just drag and drop nodes with their output supply curves and their input demand curves into a network and populate those nodes with data to shape those supply and demand curves correctly. You write systems of equations by merely dragging and dropping objects that are themselves supply and demand equations.

In our demonstration, we show the specific series of screen activities to build the model in Figure 1 (which was 8 equations and unknowns solved serially). We thereafter show the small number of screen activities to expand to the bidirectional problem in Figure 40 (which is still 8 equations and unknowns, but with twice the number of transportation links). This series of screenshots shows how simple and non-labor intensive it is to use the graphical network microeconomic equilibrium approach. This is why network microeconomists do not recoil when management or other constituents ask for substantial expansions and extensions or augmentations of your model (which they always do). This is why EIA's model" build-modify-run" cycle time can be two orders of magnitude faster. That leaves markedly more time for data assembly, market structure assembly, analysis, interactions with SME's, reporting, documentation, and communication with management and constituents. This is why EIA can always quickly adapt the model to new issues and can answer questions continuously forward in time as they emerge.

7 WHAT IF WE SOLVED THE GLOBAL WELFARE MAXIMIZATION KUHN-TUCKER CONDITIONS ANALYTICALLY? WOULD WE GET THE IDENTICAL NETWORK EQUILIBRIUM EQUATIONS?

This section has been long overdue, a long time in the preparation, and sorely needed. This section shows mathematically that if one merely groups the Kuhn-Tucker conditions derived from the monolithic global welfare maximization problem according to the individual transportation nodes, one can solve those groups **analytically** node by node and thereby dispense altogether with the need for complementarity, Kuhn-Tucker conditions, or linear programming solutions. **The Kuhn-Tucker-complementarity equations for a competitive market reduce to the network microeconomic equilibrium equations when you make the analytical substitutions.** There is absolutely no need for a numerical solution to the complementarity equations; the solution to the complementarity problem is analytical. The complementarity problem is obviated once it is analytically solved, fully replaced by the much simpler, smaller, and easier to solve network microeconomic equilibrium solution. Any time one can replace a complex, full-rank, nontransparent numerical solution with a closed-form analytical solution, doing so is a "no brainer." The analytical solution is far preferred and much simpler because you are able to use algebra and calculus (which are never wrong and never have numerical, size, or algorithm problems) rather than unnecessary numerical horsepower to solve the problem.

This section is de facto sort of a *coup de grace* for numerical complementarity or linear programming for transportation or economic modeling. The complexity and difficulty are simply not needed. **All the complexity, sophistication, and size dissolves when one makes the analytical substitutions.** EIA does not need and will not want all that complexity; EIA has had plenty of that over the years with linear programming models (which are nothing more than Leontief subproblems embedded within complementarity models as we have shown mathematically). EIA will want the simplicity of the economic solution; the simplicity is the beauty of the economic solution.

This section will show mathematically why complementarity as defined in Section 4 and precisely as defined by Gabriel op. cit. reduces exactly, precisely, and unequivocally to the network microeconomic method when one chooses to make the analytical substitutions, i.e., by being analytically opportunistic and solving problems analytically whenever one can. How many times have we seen that it is **never prudent** to leave a problem in complementarity form that can be solved exactly and precisely by analytics, i.e., to leave a solvable complementarity subproblem analytically unsolved. Every analytical solution that can be incorporated into the problem is a set of complementarity equations one need not program, need not "send to a solver," and need not solve. Substituting the analytic solution provides a direct, and as we will see colossal, reduction in the dimensionality and complexity of the transportation and economic problem EIA has to solve. We have seen previously that the number of equations and unknowns in our example drops from 32 x 32 to 8, a fourfold reduction in problem size, with the move to network microeconomic equilibrium. This section will show that you get the selfsame answer as long as you do not load a complementarity solution up with constraints or alter a global welfare function in any fashion. (If you do so, the complementarity solution becomes economically meaningless. EIA will not want to implement a global welfare maximization solution and thereby risk misuse.)

We ponder why complementarians have never to our knowledge taken the time nor effort to consider whether complementarity equations need to be solved on the computer at all, i.e., whether they are more effectively solved analytically to save time and computation? Are complementarians sufficiently enraptured with algorithms and operations research that they have not considered analytical rather than numerical solutions? Are they sufficiently eager to apply operations research algorithms or prove theorems in the literature that analytic solutions have been overlooked? We suspect so; they literally sprint to their computer. This section explores the equivalency of the methods when one does in fact opt for analytical solutions to complementarity equations. We are surprised we may be the first to actually suggest or do this.

This section shows that complementarity equations never need be used for applied economic or transportation problems. Complementarity may hold promise for monolithic constrained optimization problems that emerge in other contexts, but definitely not for spatial microeconomic problems. One simply does not need it once one solves groups of Kuhn-Tucker conditions (and therefore groups of complementarity equations) analytically rather than numerically. We get the same answer with the much simpler, direct microeconomic equations as we do with the much larger and complex complementarity equation system (as long as we have not distorted or augmented the Samuelsonian global welfare function or implemented any constraints). If we distort or augment the Samuelsonian global welfare function or implement constraints, the complementarity solution becomes nondescriptive and meaningless. Such distortion cannot occur with network microeconomic equilibrium because there is no global welfare function involved, by fiat.

We conjecture that the reason the results in this section have not been published is that the complementarity equations for problems like these have always been written in such an abstract and enumerative context that it is has been difficult to see how to group them. (The Gabriel op. cit. discussion is enumerative.) One needs the transportation network diagram in Figure 1 to identify the proper groupings of disparate complementarity equations that can be solved analytically.

Let us begin by repeating the Kuhn-Tucker conditions for the original monolithic global welfare maximization problem from Section 4. For easy identification, we have included a yellow coloring for all the Kuhn-Tucker equations that are directly related to and applicable to transportation process 1, which has an input from Hub E and has an output to hub A in our example. (The same logic applies to all other transportation processes.) We begin by rewriting the original 64 Kuhn-Tucker conditions with all the ones relevant to transportation process 1 colored in yellow.

7.1 Derivatives of Demand Equations

Here are the four derivatives of the demand equations:

$$\frac{\partial L}{\partial q_{A,17}} = -p_{17}\left(q_{A,17}\right) + \lambda_{A,17} - \mu_{A,17} = 0$$
$$\frac{\partial L}{\partial q_{B,18}} = -p_{18}\left(q_{B,18}\right) + \lambda_{B,18} - \mu_{B,18} = 0$$
$$\frac{\partial L}{\partial q_{C,19}} = -p_{19}\left(q_{C,19}\right) + \lambda_{C,19} - \mu_{C,19} = 0$$

$$\frac{\partial L}{\partial q_{D,20}} = -p_{20}\left(q_{D,20}\right) + \lambda_{D,20} - \mu_{D,20} = 0$$

7.2 Derivatives of Supply Equations

Here are the four derivatives of the supply equations:

$$\frac{\partial L}{\partial q_{21,E}} = MC_{21}\left(q_{21,E}\right) - \lambda_{21,E} - \mu_{21,E} = 0$$

$$\frac{\partial L}{\partial q_{22,F}} = MC_{22}\left(q_{22,F}\right) - \lambda_{22,F} - \mu_{22,F} = 0$$

$$\frac{\partial L}{\partial q_{23,G}} = MC_{23}\left(q_{23,G}\right) - \lambda_{23,G} - \mu_{23,G} = 0$$

$$\frac{\partial L}{\partial q_{24,H}} = MC_{24}\left(q_{24,H}\right) - \lambda_{24,H} - \mu_{24,H} = 0$$

7.3 Derivatives of the Outputs of the Transportation Activities

There are sixteen derivatives of outputs from transportation activities:

$$\frac{\partial L}{\partial q_{1,A}} = VOC_1 + \nu_{E,1} g_1{}'\left(q_{1,A}\right) - \lambda_{A,17} - \mu_{1,A} = 0$$

$$\frac{\partial L}{\partial q_{2,A}} = VOC_2 + \nu_{F,2} g_2{}'\left(q_{2,A}\right) - \lambda_{A,17} - \mu_{2,A} = 0$$

$$\frac{\partial L}{\partial q_{3,A}} = VOC_3 + \nu_{G,3} g_3{}'\left(q_{3,A}\right) - \lambda_{A,17} - \mu_{3,A} = 0$$

$$\frac{\partial L}{\partial q_{4,A}} = VOC_4 + \nu_{H,4} g_4{}'\left(q_{4,A}\right) - \lambda_{A,17} - \mu_{4,A} = 0$$

$$\frac{\partial L}{\partial q_{5,B}} = VOC_5 + \nu_{E,5} g_5{}'\left(q_{5,B}\right) - \lambda_{B,18} - \mu_{5,B} = 0$$

$$\frac{\partial L}{\partial q_{6,B}} = VOC_6 + \nu_{F,6} g_6{}'\left(q_{6,B}\right) - \lambda_{B,18} - \mu_{6,B} = 0$$

$$\frac{\partial L}{\partial q_{7,B}} = VOC_7 + \nu_{G,7} g_7{}'\left(q_{7,B}\right) - \lambda_{B,18} - \mu_{7,B} = 0$$

$$\frac{\partial L}{\partial q_{8,B}} = VOC_8 + \nu_{H,8} g_8{}'\left(q_{8,B}\right) - \lambda_{B,18} - \mu_{8,B} = 0$$

$$\frac{\partial L}{\partial q_{9,C}} = VOC_9 + \nu_{E,9} g_9{}'\left(q_{9,C}\right) - \lambda_{C,19} - \mu_{9,C} = 0$$

$$\frac{\partial L}{\partial q_{10,C}} = VOC_{10} + \nu_{F,10} g_{10}{}'\left(q_{10,C}\right) - \lambda_{C,19} - \mu_{10,C} = 0$$

$$\frac{\partial L}{\partial q_{11,C}} = VOC_{11} + \nu_{G,11} g_{11}{}'\left(q_{11,C}\right) - \lambda_{C,19} - \mu_{11,C} = 0$$

$$\frac{\partial L}{\partial q_{12,C}} = VOC_{12} + \nu_{H,12} g_{12}'(q_{12,C}) - \lambda_{C,19} - \mu_{12,C} = 0$$

$$\frac{\partial L}{\partial q_{13,D}} = VOC_{13} + \nu_{E,13} g_{13}'(q_{13,D}) - \lambda_{D,20} - \mu_{13,D} = 0$$

$$\frac{\partial L}{\partial q_{14,D}} = VOC_{14} + \nu_{F,14} g_{14}'(q_{14,D}) - \lambda_{D,20} - \mu_{14,D} = 0$$

$$\frac{\partial L}{\partial q_{15,D}} = VOC_{15} + \nu_{G,15} g_{15}'(q_{15,D}) - \lambda_{D,20} - \mu_{15,D} = 0$$

$$\frac{\partial L}{\partial q_{16,D}} = VOC_{16} + \nu_{H,16} g_{16}'(q_{16,D}) - \lambda_{D,20} - \mu_{16,D} = 0$$

7.4 Derivatives of the Inputs to the Transportation Activities

There are sixteen derivatives of inputs to transportation activities:

$$\frac{\partial L}{\partial q_{E,1}} = -\nu_{E,1} + \lambda_{21,E} - \mu_{E,1} = 0$$

$$\frac{\partial L}{\partial q_{F,2}} = -\nu_{F,2} + \lambda_{22,F} - \mu_{F,2} = 0$$

$$\frac{\partial L}{\partial q_{G,3}} = -\nu_{G,3} + \lambda_{23,G} - \mu_{G,3} = 0$$

$$\frac{\partial L}{\partial q_{H,4}} = -\nu_{H,4} + \lambda_{24,H} - \mu_{H,4} = 0$$

$$\frac{\partial L}{\partial q_{E,5}} = -\nu_{E,5} + \lambda_{21,E} - \mu_{E,5} = 0$$

$$\frac{\partial L}{\partial q_{F,6}} = -\nu_{F,6} + \lambda_{22,F} - \mu_{F,6} = 0$$

$$\frac{\partial L}{\partial q_{G,7}} = -\nu_{G,7} + \lambda_{23,G} - \mu_{G,7} = 0$$

$$\frac{\partial L}{\partial q_{H,8}} = -\nu_{H,8} + \lambda_{24,H} - \mu_{H,8} = 0$$

$$\frac{\partial L}{\partial q_{E,9}} = -\nu_{E,9} + \lambda_{21,E} - \mu_{E,9} = 0$$

$$\frac{\partial L}{\partial q_{F,10}} = -\nu_{F,10} + \lambda_{22,F} - \mu_{F,10} = 0$$

$$\frac{\partial L}{\partial q_{G,11}} = -\nu_{G,11} + \lambda_{23,G} - \mu_{G,11} = 0$$

$$\frac{\partial L}{\partial q_{H,12}} = -\nu_{H,12} + \lambda_{24,H} - \mu_{H,12} = 0$$

$$\frac{\partial L}{\partial q_{E,13}} = -\nu_{E,13} + \lambda_{21,E} - \mu_{E,13} = 0$$

$$\frac{\partial L}{\partial q_{F,14}} = -\nu_{F,14} + \lambda_{22,F} - \mu_{F,14} = 0$$
$$\frac{\partial L}{\partial q_{G,15}} = -\nu_{G,15} + \lambda_{23,G} - \mu_{G,15} = 0$$
$$\frac{\partial L}{\partial q_{H,16}} = -\nu_{H,16} + \lambda_{24,H} - \mu_{H,16} = 0$$

7.5 Derivatives of the Lagrangian With Respect to the Lagrange Multipliers λ

The balance equations are the derivative of the Lagrangian with respect to the Lagrange multipliers λ:

$$\frac{\partial L}{\partial \lambda_{A,17}} = q_{A,17} - q_{1,A} - q_{2,A} - q_{3,A} - q_{4,A} = 0$$
$$\frac{\partial L}{\partial \lambda_{B,18}} = q_{B,18} - q_{5,B} - q_{6,B} - q_{7,B} - q_{8,B} = 0$$
$$\frac{\partial L}{\partial \lambda_{C,19}} = q_{C,19} - q_{9,C} - q_{10,C} - q_{11,C} - q_{12,C} = 0$$
$$\frac{\partial L}{\partial \lambda_{D,20}} = q_{D,20} - q_{13,D} - q_{14,D} - q_{15,D} - q_{16,D} = 0$$
$$\frac{\partial L}{\partial \lambda_{21,E}} = q_{E,1} + q_{E,5} + q_{E,9} + q_{E,13} - q_{21,E} = 0$$
$$\frac{\partial L}{\partial \lambda_{22,F}} = q_{F,2} + q_{F,6} + q_{F,10} + q_{F,14} - q_{22,F} = 0$$
$$\frac{\partial L}{\partial \lambda_{23,G}} = q_{G,3} + q_{G,7} + q_{G,11} + q_{G,15} - q_{23,G} = 0$$
$$\frac{\partial L}{\partial \lambda_{24,H}} = q_{H,4} + q_{H,8} + q_{H,12} + q_{H,16} - q_{24,H} = 0$$

7.6 Derivatives of the Lagrangian with Respect to the Lagrange Multipliers ν

The input-output equations are the derivative of the Lagrangian with respect to the Lagrange multipliers ν:

$$\frac{\partial L}{\partial \nu_{E,1}} = g_1\left(q_{1,A}\right) - q_{E,1} = 0$$
$$\frac{\partial L}{\partial \nu_{F,2}} = g_2\left(q_{2,A}\right) - q_{F,2} = 0$$
$$\frac{\partial L}{\partial \nu_{G,3}} = g_3\left(q_{3,A}\right) - q_{G,3} = 0$$

$$\frac{\partial L}{\partial \nu_{H,4}} = g_4\left(q_{4,A}\right) - q_{H,4} = 0$$

$$\frac{\partial L}{\partial \nu_{E,5}} = g_5\left(q_{5,B}\right) - q_{E,5} = 0$$

$$\frac{\partial L}{\partial \nu_{F,6}} = g_6\left(q_{6,B}\right) - q_{F,6} = 0$$

$$\frac{\partial L}{\partial \nu_{G,7}} = g_7\left(q_{7,B}\right) - q_{G,7} = 0$$

$$\frac{\partial L}{\partial \nu_{H,8}} = g_8\left(q_{8,B}\right) - q_{H,8} = 0$$

$$\frac{\partial L}{\partial \nu_{E,9}} = g_9\left(q_{9,C}\right) - q_{E,9} = 0$$

$$\frac{\partial L}{\partial \nu_{F,10}} = g_{10}\left(q_{10,C}\right) - q_{F,10} = 0$$

$$\frac{\partial L}{\partial \nu_{G,11}} = g_{11}\left(q_{11,C}\right) - q_{G,11} = 0$$

$$\frac{\partial L}{\partial \nu_{H,12}} = g_{12}\left(q_{12,C}\right) - q_{H,12} = 0$$

$$\frac{\partial L}{\partial \nu_{E,13}} = g_{13}\left(q_{13,D}\right) - q_{E,13} = 0$$

$$\frac{\partial L}{\partial \nu_{F,14}} = g_{14}\left(q_{14,D}\right) - q_{F,14} = 0$$

$$\frac{\partial L}{\partial \nu_{G,15}} = g_{15}\left(q_{15,D}\right) - q_{G,15} = 0$$

$$\frac{\partial L}{\partial \nu_{H,16}} = g_{16}\left(q_{16,D}\right) - q_{H,16} = 0$$

7.7 Accumulating and Solving the Equations for Transportation Process 1

We now assemble the Kuhn-Tucker conditions that are associated with transportation process 1, the yellow equations, in a single location:

$$\frac{\partial L}{\partial q_{1,A}} = VOC_1 + \nu_{E,1} g_1{}'\left(q_{1,A}\right) - \lambda_{A,17} - \mu_{1,A} = 0$$

$$\frac{\partial L}{\partial q_{E,1}} = -\nu_{E,1} + \lambda_{21,E} - \mu_{E,1} = 0$$

$$\frac{\partial L}{\partial \nu_{E,1}} = g_1\left(q_{1,A}\right) - q_{E,1} = 0$$

$$q_{1,A}\mu_{1,A} = 0 \quad \mu_{1,A} \geq 0 \quad q_{1,A} \geq 0$$

$$q_{E,1}\mu_{E,1} = 0 \quad \mu_{E,1} \geq 0 \quad q_{E,1} \geq 0$$

There is no reason to leave these five equations as Kuhn-Tucker conditions or to convert them to complementarity types of equations. Absolutely the contrary; it is going to prove far more prudent to solve them analytically. For any bona fide production function or input-output function one specifies, we would want to solve these equations analytically if at all possible and avoid the more difficult setting. We would want the analytical solution to these equations to be used in place of the complementarity equations because the analytic solution is much simpler, more direct, more insightful, less size- and computationally-intensive, and less prone to implementation or numerical error. We would want to compress these difficult equations into very simple equations that comprise their solution. To illustrate how, supposed we used the production function from Figure 14, which was:[31]

$$x = y^a = g\left(y\right) \Rightarrow g'\left(y\right) = ay^{a-1}$$

(Even with complementarity, one invariably has to appeal to a specific input-output function before one can implement a model, so this is no “stretch.) We substitute the input-output function and the derivative of the input-output function for transportation process 1:

$$VOC_1 + \nu_{E,1} a q_{1,A}{}^{a-1} - \lambda_{A,17} - \mu_{1,A} = 0$$
$$\nu_{E,1} = \lambda_{21,E} - \mu_{E,1} = 0$$
$$q_{1,A}{}^{a} - q_{E,1} = 0$$
$$q_{1,A}\mu_{1,A} = 0 \quad \mu_{1,A} \geq 0 \quad q_{1,A} \geq 0$$
$$q_{E,1}\mu_{E,1} = 0 \quad \mu_{E,1} \geq 0 \quad q_{E,1} \geq 0$$

We eliminate the ν term by substituting the first two equations:

$$VOC_1 + \left(\lambda_{21,E} - \mu_{E,1}\right) a q_{1,A}{}^{a-1} - \lambda_{A,17} - \mu_{1,A} = 0$$
$$q_{1,A}{}^{a} - q_{E,1} = 0$$
$$q_{1,A}\mu_{1,A} = 0 \quad \mu_{1,A} \geq 0 \quad q_{1,A} \geq 0$$

[31] The generality of this development is not impeded by selecting a specific input-output function. All input-output functions that would realistically represent a transportation or economic process apply identically.

$$q_{E,1}\mu_{E,1}=0 \quad \mu_{E,1}\geq 0 \quad q_{E,1}\geq 0$$

We now have to analyze two cases for q_{1A}: (1.) Case 1: $q_{1A}>0$ and (2.) Case 2: $\mu_{1A}>0$. We will analyze each now.

Case 1: $q_{1A}>0$

If $q_{1A}>0$, the fourth Kuhn-Tucker condition implies that $\mu_{1A}=0$. Furthermore, the second Kuhn-Tucker condition implies that $q_{E1}>0$. That in turn implies that $\mu_{E1}=0$. Substituting those conditions into the Kuhn-Tucker conditions gives the relationships:

$$VOC_1+\lambda_{21,E}aq_{1,A}{}^{a-1}-\lambda_{A,17}=0$$
$$q_{1,A}{}^{a}-q_{E,1}=0$$

These two equations are very simple to solve analytically by substitution. And when we solve them, we have the full analytical solution to the five yellow Kuhn-Tucker conditions for Case 1 for transportation process 1, and therefore need not consider those five Kuhn-Tucker conditions or their complementarity equivalent any further. We are finished with them; they are fully solved as follows:

$$q_{1,A}=\left(\frac{\lambda_{A,17}-VOC_1}{a\lambda_{21,E}}\right)^{\frac{1}{a-1}}$$

$$q_{E,1}=q_{1,A}{}^{a}=\left(\frac{\lambda_{A,17}-VOC_1}{a\lambda_{21,E}}\right)^{\frac{a}{a-1}}$$

This solution would ONLY exist when:

$$\lambda_{A,17}-VOC_1>0$$

Ever seen these equations before? **<u>You BET we have.</u>** They are the first of the two profit maximization equations for transportation process indexed 1 derived in Section 5 related to network equilibrium! **<u>We have just analytically derived the profit maximizing equations for microeconomic network equilibrium, directly from the complementarity equations, reducing their detail and complexity substantially</u>**. This is profound; these are the selfsame equations derived directly from producer profit maximization in Section 5, which are rewritten here for convenient reference:

$$x^*(p,w)=\begin{cases}\left(\dfrac{p-\phi}{aw}\right)^{\frac{a}{a-1}} & \text{if } p-\phi>0\\ 0 & \text{if } p-\phi\le 0\end{cases}$$

$$y^*(p,w)=\begin{cases}\left(\dfrac{p-\phi}{aw}\right)^{\frac{1}{a-1}} & \text{if } p-\phi>0\\ 0 & \text{if } p-\phi\le 0\end{cases}$$

Case 2: $\mu_{1A}>0$.

In this situation, we know that $q_{1A}=0$. We know from the second Kuhn-Tucker condition that $q_{E1}=0$. Substituting into the Kuhn-Tucker conditions, we have:

$$\begin{gathered}VOC_1-\lambda_{A,17}=\mu_{1,A}\\ q_{1,A}\mu_{1,A}=0 \quad \mu_{1,A}>0 \quad q_{1,A}=0\\ 0\mu_{E,1}=0 \quad \mu_{E,1}\ge 0 \quad q_{E,1}=0\end{gathered}$$

We can manipulate this:

$$\begin{aligned}&VOC_1-\lambda_{A,17}=\mu_{1,A}>0\Rightarrow\lambda_{A,17}-VOC_1<0\\ &q_{1,A}=q_{E,1}=0\end{aligned}$$

This is the second half of the Kuhn-Tucker conditions above.

Ponder what we have just proven. We have proven that in fact we **<u>can</u>** solve the Kuhn-Tucker conditions related to transportation process 1 **<u>analytically</u>**. When we do so, we get the solution in Section 5, the standard microeconomic solution. The same is true for all of the other transportation links in the system and in fact for every economic node in the network that comprises your model. There is absolutely no need to solve the ponderous 64 equation system as a complementarity problem or even its smaller 32 equation cousin. In fact to do so would be a waste of time. If you can get a reliable solution analytically, you are imprudent to rely instead on a costly, wasteful, opaque, monolithic, 1970s-computer-architecture numerical or algorithmic procedure that is oversized by a factor of four in terms of numbers of equations. Analytical solutions trump numerical solution. Analytical solutions do not consume time, resources, size, computer cycles, or numerical accuracy. To rely on a numerical solution when you have an analytical solution is not what you want to do. You are able to eliminate these Kuhn-Tucker-complementarity conditions and go with the analytical solution. We have just proven it. Complementarians can be happy; their problem has been solved analytically.

It is useful to summarize our findings here:

$$q_{E,1} = \begin{cases} \left(\dfrac{\lambda_{A,17} - VOC_1}{a\lambda_{21,E}} \right)^{\frac{a}{a-1}} & \text{if } \lambda_{A,17} - VOC_1 \geq 0 \\ 0 & \text{if } \lambda_{A,17} - VOC_1 < 0 \end{cases}$$

$$q_{1,A} = \begin{cases} \left(\dfrac{\lambda_{A,17} - VOC_1}{a\lambda_{21,E}} \right)^{\frac{1}{a-1}} & \text{if } \lambda_{A,17} - VOC_1 \geq 0 \\ 0 & \text{if } \lambda_{A,17} - VOC_1 < 0 \end{cases}$$

These are precisely the previously solved Kuhn-Tucker conditions in the network microeconomic equilibrium analysis in Section 5. Once you select an input-output function, there is no need to solve the Kuhn-Tucker conditions, which are the complementarity conditions, using any complex numerical algorithm. Not only is there no need; in fact, it would be extremely imprudent and wasteful. You have already solved them; the need is fully met and fully complete. You have just solved analytically what complementarians insist on solving numerically. Analytics trumps numerics every time.

Observing this, complementarity is best focused on complex constrained optimization problems (other than market problems) with a single monolithic decision maker and a complex constraint set comprised of equalities and inequalities. A refinery problem (which has traditionally used linear programming) would be a prime example, an ideal venue to deploy complementarity. Complementarity would seem to be useful when you have a complex constraint set. It is a distant second best for economic and market problems. If one insisted on complementarity, one can substitute the functional form for the transportation nodes and save all the trouble and all the algorithm size, complexity, computation time, indirectness, "sending things to a solver," and programming.

It is useful to note that the five Kuhn-Tucker-complementarity equations that we began with have reduced to two very simple equations, a 60 percent reduction in the number of equations and unknowns that will have to be solved. Reduction of the Kuhn-Tucker conditions to an analytical form has cut the problem size by more than half, not to mention eliminating difficult numerical procedures in favor of easily executable analytical calculations. Furthermore, the two equations that remain (a supply equation and a demand equation) are simple and direct and importantly are no longer in the form of a complementarity equation. No more need for complementarity; complementarity has done its job. It has given you a set of equations to solve analytically. Once you solve a complementarity problem analytically, the solution is no longer in complementarity form. **Simple substitution of the Kuhn-Tucker conditions eliminates altogether the need for a complementarity solution for transportation (and in fact for any producer)**. The entire solution is analytical. You don't need complementarity; you have the solution analytically derived from it.

The implications of this finding are rather far reaching. You are certainly free to formulate your problem as a dramatically oversized, ponderous, monolithic global welfare maximization problem, and write your 64 Kuhn-Tucker conditions and your equivalent complementarity equations in all their splendorous detail and then try to program them, in spite of the shortcomings and risks. However, we have just proven that if you posit production functions or input-output relationships, **<u>you can solve all the complementarity equations analytically rather than numerically and dramatically reduce the number and complexity of your equation system</u>**. You have no need whatsoever for a numerical complementarity algorithm. When you obtain and use the analytical solution, you end up precisely with the network microeconomic equations and solution, which you can just solve directly as we have proven. Complementarity has merely been a formulation procedure. You have "analyzed" your way out of having to deploy complementarity algorithms, which are dramatically oversized. You avoid the complexity and huge dimensionality of those solution methods. People are ill advised to spend the time and complexity solving numerically what you have already solved analytically. You are better advised to use the network microeconomic approach directly and save the hassle, computer time, programming time, and analyst time by solving complementarity analytically, bundling them into objects, and dragging those objects onto a page in the form of a network.

Someone may say *"Yeah, sure, it may be true for **<u>that</u>** input-output function, but it doesn't work for others."* However, such an assertion is incorrect. The developments in this and previous sections work for every input-output function that would realistically characterize transportation (or any other economic activity)[32]. There are no limitations because of the specific production function we have chosen for illustrative purposes (other than the desire for nonincreasing returns to scale). The equivalency in this section is largely independent of the production function analytical form we use. Sure, the resultant supply and demand equations will be structurally and parametrically different to a degree with different input-output functions. But the proofs all go through to the same conclusion in exactly the same way. The proof that analytical solution of complementarity is always possible and reduced the size and complexity of complementarity by at least a factor of four is invariant with respect to the specific production function chosen.

Think of what we have proven. In a transportation economic equilibrium problem, **<u>we neither need nor want complementarity. If we select complementarity as a formulation technique, we solve it analytically, never numerically, and reduce it to the much simpler network microeconomic equilibrium problem by so doing!</u>** Network equilibrium gives the same answer for the example problem using a much simpler and more direct calculation. Complementarians are happy; they can cheer the complementarity formulation. Economists are happy; can cheer the decentralized agent solution. Management will be happy; they get the answers in two orders of magnitude less time and without analytical or error risk.

We believe the reason that complementarians have not considered an analytical solution for the economic problem is that, in the broad the set of problems they approach outside the economic realm, rarely if ever is there an analytical solution. To search for one is futile. Large optimal control or optimization problems replete with myriad constraints are simply not going to admit of an analytical solution. However, the competitive economic situation is unique—it DOES admit of an analytical solution. It is because of the specific structure of the complementarity equations in the

[32] We will want the production functions to display non-increasing returns to scale. This is true for both monolithic optimization and network microeconomic equilibrium. Increasing returns to scale production functions introduce dimensions of complexity we do not want to address here. That is for a later date.

competitive economic setting that allows an analytical solution of the complementarity equations. This paper does complementarians a great service, showing them how to solve the problem they pose analytically rather than numerically. Numerical complementarity holds promise for nonlinear optimization in applications outside economics. However, for nonlinear economics, the solutions to nonlinear complementarity problems are analytical, and that eliminates the need for numerical solutions.

Complementarians cannot have it both ways. If they insist on asserting that the complementarity solution is the same as network equilibrium (equivalent method, same answer), then assuredly the analytical solution we have developed is available to them and far and away better than the numerical solution toward which they are otherwise so eager to sprint. When they solve their complementarity equations analytically, they get identically the microeconomic network equilibrium equations, absolutely identically, which are much simpler and smaller to solve. If they argue that their solution is the same as the network solution, then the smaller solution is very much better for them. Since the network solution is 1/8 of the size and complexity, assuredly they will want to use the network solution to solve their complementarity equations. There is simply never a case to be made for complementarity solution algorithms or implementation in distributed economic problems.

8 HUB-AND-SPOKE TRANSPORTATION CONTRASTED WITH POINT-TO-POINT TRANSPORTATION

EIA is aware that where transportation and regionality are concerned, aggregation is an insurmountable difficulty that can doom your efforts. You need the maximum degree of regional and interconnecting transportation detail you can possibly get. You do not want to be forced to aggregate just because your implementation method is too ponderous or too limited to allow tens of thousands of transportation links over dozens of time points. Transportation is not a problem that is amenable to aggregation or approximation. You need the spatial detail.

There are several other dimensions of transportation of which EIA needs to be cognizant. This section summarizes those other dimensions and offers solutions.

8.1 Hub-and-Spoke Transportation

In the 1980s, airlines and shipping companies discovered the notion of "hub-and-spoke" transportation.[33] EIA's transportation and logistical modeling system must be able to represent hub-and-spoke transportation as well as point-to-point transportation. (The example in the previous sections focused on point-to-point transportation, but as we shall see here, that does not compromise their generality.)

Hub-and-spoke transportation would occur as follows in world LNG markets. There would be a "Strait of Hormuz" hub. All LNG tankers entering or exiting the Persian Gulf would de facto rally and compete at the Strait of Hormuz hub. There would be de facto FOB against FOB competition at the Strait of Hormuz hub. Then the ships would leave the Strait of Hormuz hub and travel to various other hubs—a Pacific Coast North America hub (Northwest Pacific), a Northwest Pacific (offshore Asia) hub, a Southwest Pacific (Southeast Asia offshore) hub, a Cape of Good Hope hub, etc. After rallying and competing on a FOB against FOB basis at the Northwest Pacific hub, the ships would enter ports in Korea, Japan, or China from there. The diagram in Figure 41 illustrates.

Contract flows are de facto point-to-point flows, although not necessarily. Spot flows are hub-and-spoke flows. Tankers would be de facto competing on an FOB[34] against FOB basis everywhere in the world, seeing the best price for their cargo they can get and exchanging cargos and routes at the drop of a hat.

8.2 A Sample Hub-and-Spoke Transportation Problem

A more general, more obvious hub-and-spoke transportation system appears as in Figure 42. Notice that there are three hubs (designated as hubs in the diagram), and there are three spokes or feeder systems surrounding each of the three hubs. If one were to attempt to implement this as a complementarity problem, let us count the hubs and activities to see how large it is. There are:

- 12 hubs (circular nodes) in total.

[33] Our company Decision Focus Incorporated was closely involved in the transportation and revenue management revolution of the 1980s.
[34] FOB is an abbreviation for "freight on board," referring to cargo on a ship.

- 24 transportation processes in total.
- 9 demand processes in total.
- 9 supply processes in total.

Figure 41: A Hub-and-Spoke Representation of World LNG Tanker Transportation

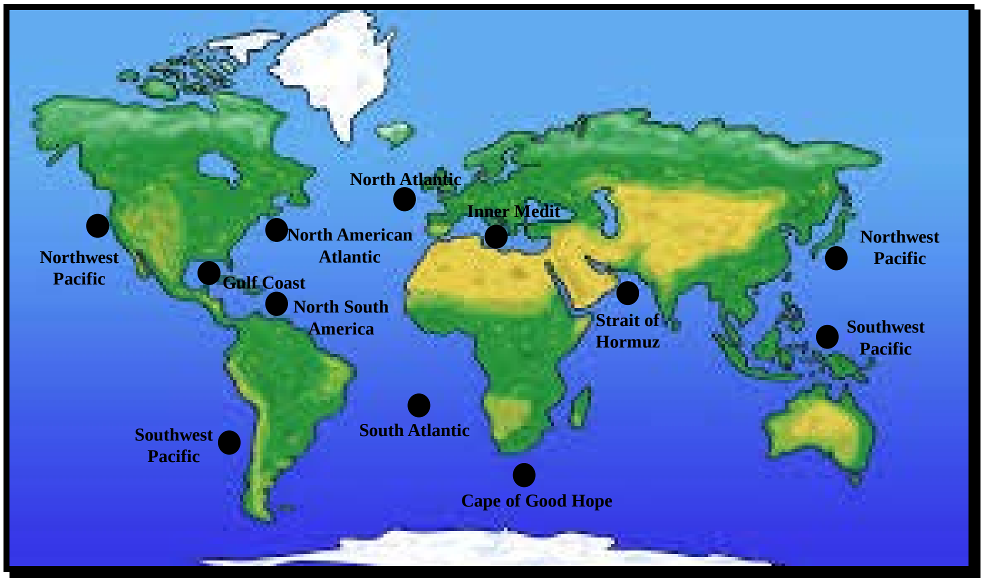
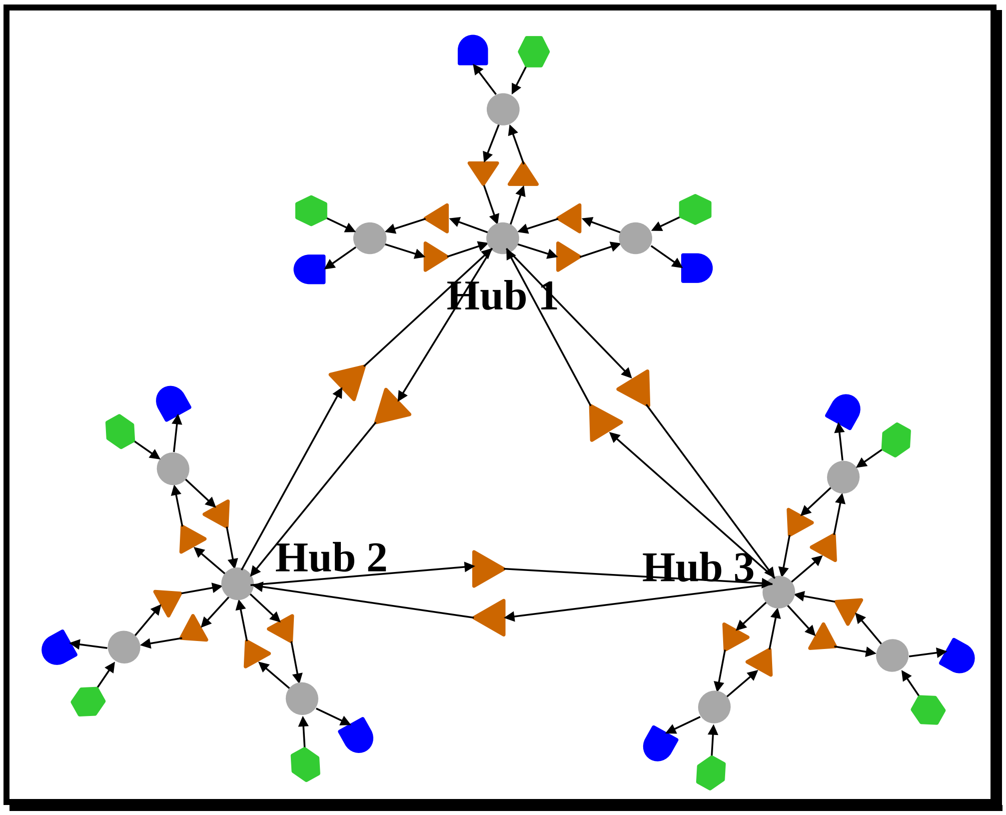


Figure 42: Typical Hub-and-Spoke Transportation System



That means that there are 24+18 = 42 economic nodes in the hub-and-spoke network, and there are 12 hub nodes in the model. More importantly, there are 66 links in the model, and each link contains an (unknown) flowing quantity. It is the number of links that determines the number of unknown quantities. Thus, there are 12 unknown prices and 66 unknown quantities in this hub-and-spoke model. The Jacobean matrix that would be needed to characterize and solve the complementarity equations would be of rank equal to the sum of unknown prices and unknown quantities, i.e., 12+66=78. The Newton's method required to solve this problem would require creating on every iteration and inverting on every iteration a 78 x 78 extremely sparse matrix. You would have to write 78 complementarity or equality equations and differentiate every one with respect to every one of the 78 variables and insert them properly into a 78 x 78 Jacobian matrix, and you would have to do that real-time on every Newton iteration. (And there would be hundreds or thousands of Newton iterations, none of which is guaranteed to converge to a solution according to Gabriel.) If the structure of the problem changed, you would have to do it all over again. Then you would have to invert that matrix in order to take the next step in the iterative algorithm. You have no guarantee of convergence. This is a very large and risky formulation of the problem indeed. We surmise there simply aren't enough grad students to be coerced to do all this work!

In sharp contrast, how many aggregate supply-demand curve pairs are there in this model? **There are only 12, one for each circular hub node.** There are a mere 12 supply-demand curve pairs here! The outputs of all economic nodes have a supply curve; that is a direct and easy byproduct of network microeconomic methods. The inputs to all economic nodes have a demand curve. Therefore, all the input links to all the hubs carry a supply curve. All the output links from all the hubs carry a demand curve. This is actually a very simple problem for network equilibrium. You write zero programs. None! You simply populate the transportation, supply, and demand nodes with data and find the simultaneous clearing of the twelve supply-demand curve pairs. This is quite a simple problem, and the dimensionality is so much lower. Assuming the input-output functions were all programmed as articulated in Section 5, no one would approach the problem in any way other than network microeconomic equilibrium—zero programming, very simple algorithm, very small dimensionality of solution method, entirely local solution method with no Hessians or Jacobians or other large matrices, no risk of polluting the solution by objective function alteration or constraint addition. The hub-and-spoke network would be a difficult and risky problem for complementarians to implement. Readers should give it a try by attempting to write the equations in the manner we used in Section 4. Have plenty of paper handy!

This is fairly compelling. The hub-and-spoke problem would require a 78 x 78 solution method if complementarity were used. It would require a solution of 12 supply-demand curve pairs if network microeconomics were used. That is a factor of 6 smaller, and certainly a factor of 100 lower probability of error and time commitment.

8.3 Hybrid Hub-and-Spoke and Point-to-point Transportation Systems

What if you had to represent a point-to-point transportation system embedded inside this hub-and-spoke transportation system? "Are you crazy," one might think. "Isn't it one or the other?" Absolutely the contrary. The reality is that LNG tanker transportation around the world is a **composite** of point-to-point transportation for contracted flows and hub-and-spoke transportation for spot flows. To a network microeconomic modeler, dragging in the hub-and-spoke structure as in Figure 42 onto the page and dragging in all the contract path point-to-point transportation processes is extremely straightforward. It does not involve any equation writing, coding, or programming and can be done in an hour or less.

EIA needs a transportation modeling methodology that allows hub-and-spoke, point-to-point, and in fact a hybrid between the two. Network microeconomic equilibrium allows that to be done straightforwardly and quickly. This section has shown how.

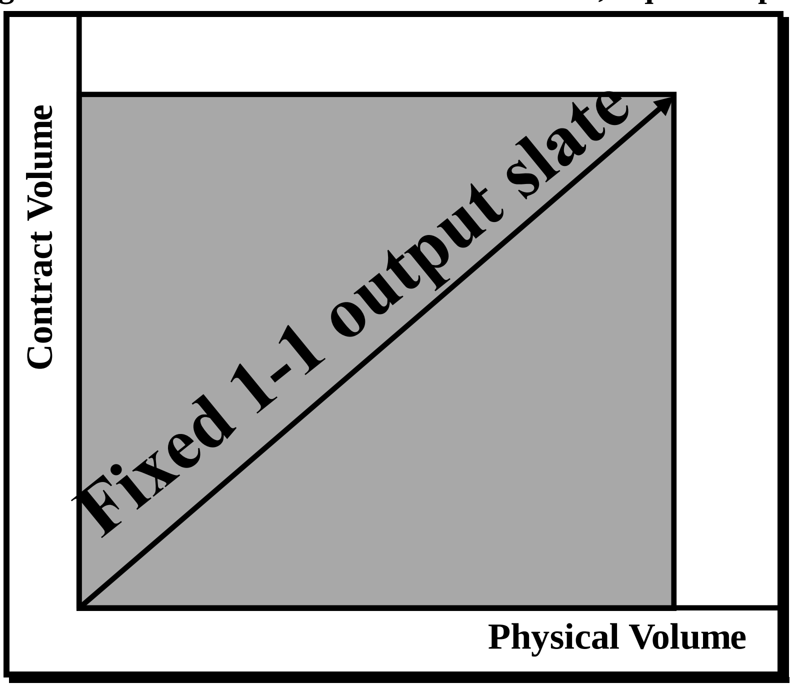
9 CONTRACTED VERSUS SPOT TRANSPORTATION (AND SUPPLY)

One of the phenomena that has been strong and persistent in LNG and pipeline transportation has been contracting, particularly take or pay contracting. This is such a strong force and has such a strong impact on prices, quantities, regional disposition of supply, and other market forces that it must be explicitly represented in any EIA world transportation model. This section articulates how it can be very easily done using network equilibrium methods. We wouldn't know how to approach the problem using complementarity, and we are not motivated to find out.

The notion behind a take or pay contract is just that—you pay for the gas whether you take it or not. A take or pay source of gas is the **most** inframarginal source in your supply portfolio. It is the first in/last out supply in your portfolio because you have a contractual obligation to take it. How can we model that?

Suppose we had an economic process that had two outputs, and those two outputs were produced in exactly equal, fixed, immutable proportions, i.e., one to one. Suppose the first output represented physical gas and the second output represented contract requirement or contracted gas. The production function for such a process appears in Figure 43. With this production function, you produce one unit of contracted gas for every unit of physical gas, i.e., they produce as a matched pair. You satisfy one unit of contract volume commitment with every unit of physical gas delivered, one for one.

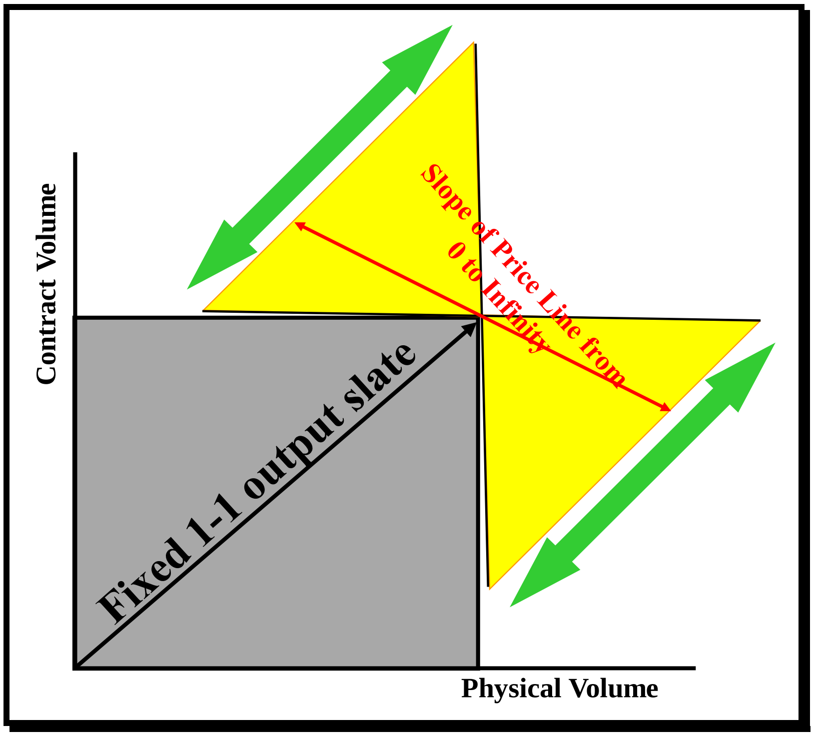
Figure 43: Production Function with Fixed, Equal Outputs



If we had a joint production function in the form in in Figure 43, we would get the same 1-1 ratio of contract gas and physical gas no matter what the relative price of contract gas and physical gas. Figure 44 illustrates that you get a 1-1 ratio no matter what the relative prices. The relative prices are indicated by the slope of the tangent line to the production function surface.

In this situation, every price line (i.e., every price ratio) leads to the same operating point between physical gas and contract gas. There is obviously a unique quantity ratio but a completely non-unique price. Every price leads to the same quantity. There is an infinity of prices but a single quantity ratio.

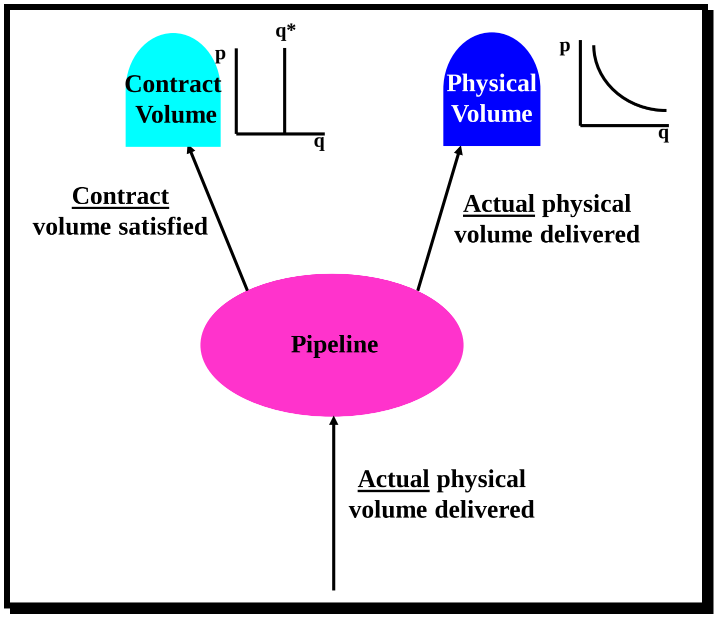
Figure 44: Contract and Physical Gas Are Produced Equally with Fixed, Equal Outputs No Matter What the Prices



Suppose we crafted a pipeline network as shown in Figure 45. We posit an absolutely fixed contract volume in the form of a fixed, inelastic demand for contract gas at the contract volume q*. That is represented using the inelastic demand curve on the contract gas output from the node in Figure 45. This structure will mandate that the contract demand q* is satisfied by the dual output node. There will absolutely have to be q* units of contract gas flowing to the inelastic contract demand node.

What does that mean for the physical gas node? It means, by virtue of the 1-1 relationship, that q* units of physical gas will be forced into the physical gas demand curve at the right. That is precisely what a contract does. It mandates delivery of physical gas in the amount q* to the physical market in the demand node at the right. That is precisely what we want—contracted delivery in the amount q* forced into the physical market at the right no matter what the price. The price in the physical market is irrelevant. We want to force the contracted quantity q* into the physical market, and this construct does exactly that.

Figure 45: Inelastic Demand for Contract Gas



The price of physical gas will be the market price in the upstream and downstream region. Generally the price of physical gas at the delivery point will be lower than the price of gas at the upstream point plus the cost of the pipe plus any losses. Therefore, the price in the contract market will be equal to the loss in the market that occurs because of the contract that forces delivery. The price in the contract market will be how far "out of the money" as measured by the market price the contract is. This is a keenly interesting number, for it tells us not only the price in the upstream and downstream physical market, but it also tells us how much loss the market would incur because they contracted a certain volume of gas with a price that is lower than their acquisition price plus transportation cost. This is a very convenient structure, for it calculates how far out of the money or how far in the money the take or pay contract is. It de facto marks the contract to market at the observed prices in the market for product and for input factor.

Let us now add one more wrinkle. The contract volume must be a "greater than or equal to" volume. The market at the point of delivery must take at least the volume q*, and it may take more if it wants. That situation, which is the real world situation, is very simple to implement in this network framework. All one must do is implement a "free disposal" demand curve in the contract market. In such a situation, if the physical market wants to take a quantity q larger than q*, there must be zero contract value placed on the volume q - q*. That is accomplished by implementing the structure in Figure 46. In that structure, the free disposal demand curve has the structure in Figure 47. Any amount of quantity will be consumed at a price of $\$1x10^{-8}$/MMBtu or lower. If the physical market wants to exceed the contract commitment, the difference between the physical market and the contract market will be made up at an infinitesimal price of $\$1x10^{-8}$ per MMBtu or lower. This is rather an ingenious construct, forcing at least the contract quantity q* into the market at the point of delivery while at the same time allowing more than the contract quantity q* to be purchased at the point of delivery if that additional gas is competitive in the market at the point of delivery without any subsidy or contract obligation. This is actually quite a sophisticated model of gas take or pay contracts, easily available with the network modeling structure summarized and applied here. We doubt it is available with other solution methods, and it is certainly not available as easily, straightforwardly, or transparently.

Figure 46: Free Disposal Means Physical Volume ≥ Contract Volume

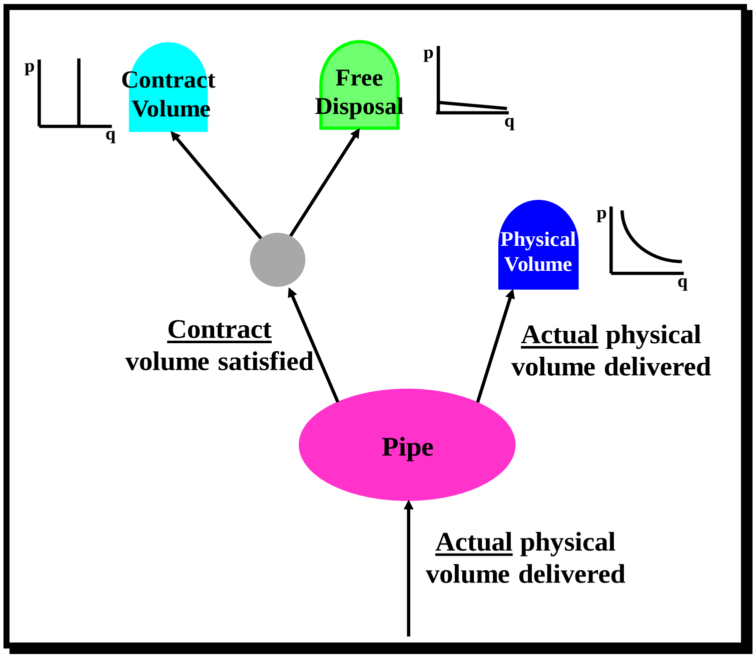
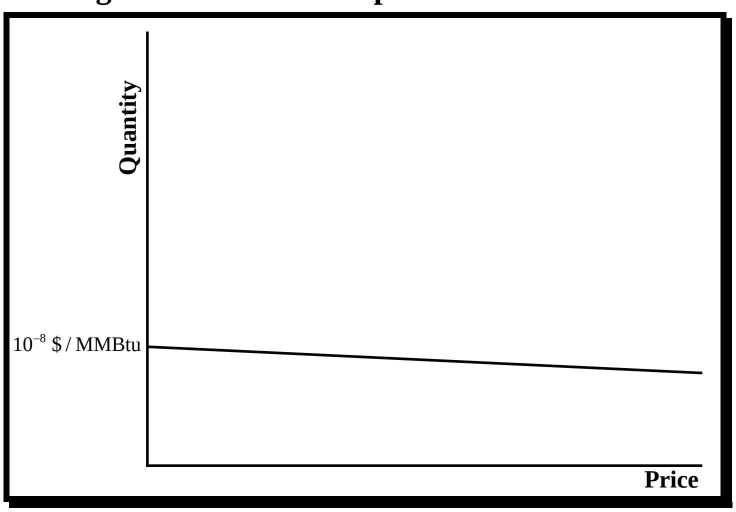


Figure 47: Free Disposal Demand Curve



It can be implemented on any transportation link in the world, allowing EIA to have a hybrid of take or pay contracts and no contracts (spot deliveries). This is not only extremely powerful, but it is representative of the way many LNG project and pipelines in the world have worked. It can also help you analyze such questions as "What happens if KOGAS breaks its contract with the Northwest Shelf," a phenomenon that has occurred. See

https://au.news.yahoo.com/thewest/a/17368370/gorgon-gas-sales-fall-short/

and

http://www.industrialinfo.com/news/abstract.jsp?newsitemID=235731.

What are the implications of this to the supplier? What are the implications of this to the customer? What are the implications of this for world spot price and contract deliveries? These are the questions that network structure and logic such as the foregoing can answer, and EIA needs to represent. We would challenge other modeling structures ever to do this correctly unconditionally. Keep firmly in mind, this structure forces the take or pay volume into the physical market. If the market is over-contracted, the physical gas has to be liquidated at low spot price in the target spot market, exactly the way the real world works. One need only look to environments like Italy or Germany to note that over-contracted Russian or Algerian gas has to be liquidated in local spot markets, and the party holding the take or pay contract takes a significant "bath" when that occurs. The magnitude of the "bath" he or she takes will be the price times quantity that shows up in this modeling construct at the inelastic contract demand node. That is how much money is lost by the contract if it is ever too large on the quantity side or ever out of the money on the price side.

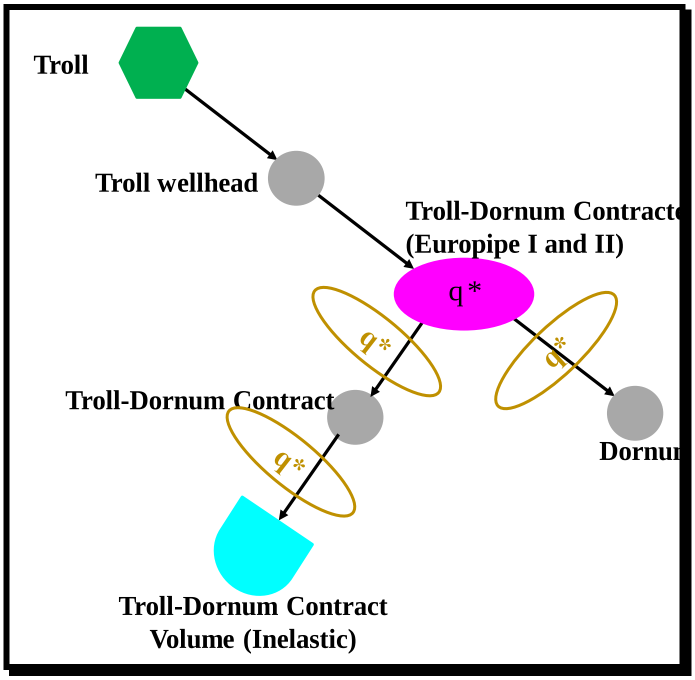
Note that in order to build this general construct, you need continuous demand functions for all products with prices that range from zero to infinity and quantities that range from zero to infinity. That requirement renders it difficult to impossible for linear programming to work in this environment. We have shown before (and EIA and FEA have known since the days of PIES that, in fact, linear programming has a "discrete countable price" problem and therefore requires inelastic demand). We must have a very flat demand curve at a very low price in order to implement the "free disposal" property, meaning that at a nearly zero price, the quantity consumed can become infinite.[35] We can craft that demand function easily in network equilibrium models and insert it where it needs to be. We must also craft a completely inelastic demand curve for the contract market, which is easy for both monolithic global welfare maximization and network equilibrium.

Let us now move from the modeling world to the real world. The Troll field is a large gas field in the North Sea owned and controlled by Norway. There is an undersea pipeline system from the North Sea that lands in various points in Northwest Europe, one of those points being Dornum in Northwestern Germany. There are two pipes, Europipe I and Europipe II that emanate in the Norwegian North Sea and terminate at Dornum, Germany. At present, the customers in Dornum, Germany have a take or pay contract for deliveries through Europipe I and Europipe II. They are de facto forced to take deliveries at Dornum from these two pipes at the level of the contracted volumes in their contracts with the Norwegian pipe. What if we had a modeling construct that forced supply into Dornum at precisely the contract volume q* at Dornum, but it allowed customers to take spot gas through the pipe at volumes larger than the contract volume? Wouldn't that represent the take or pay contract from the Germans at Dornum to the Norwegians at the upstream end of the two Europipes? The answer is a definitive **yes**! That would be precisely the

[35] No one has proven it yet to our knowledge, but we suspect that nonlinear monolithic optimization has precisely the same bang-bang, countably discrete price problem. We will not endeavor to prove it, but it seems intuitively true. The generality of nonlinear rather than linear formulations makes it difficult to prove in general, but consider that "almost linear" problems certainly would have the countable price problem. In our view, this is too big a risk to take to allow complementarity algorithms to ever be used.

right structure to represent take or pay contracts at Dornum, which are nothing more than economic instruments that guarantee that the contract path is the **most inframarginal source** into the market at Dornum. That is what a take or pay contract is—it guarantees that the volumes under the take or pay contract are the most inframarginal of all the sources into the market. (All the financial guarantees that secure that contract occur outside the market.) With the foregoing construct, we could consider the network diagram in Figure 48, which would ensure forced entry of exactly the contracted quantity q* into the market at Dornum.

Figure 48: Forced Contract Entry of Exactly q* into Dornum from Troll

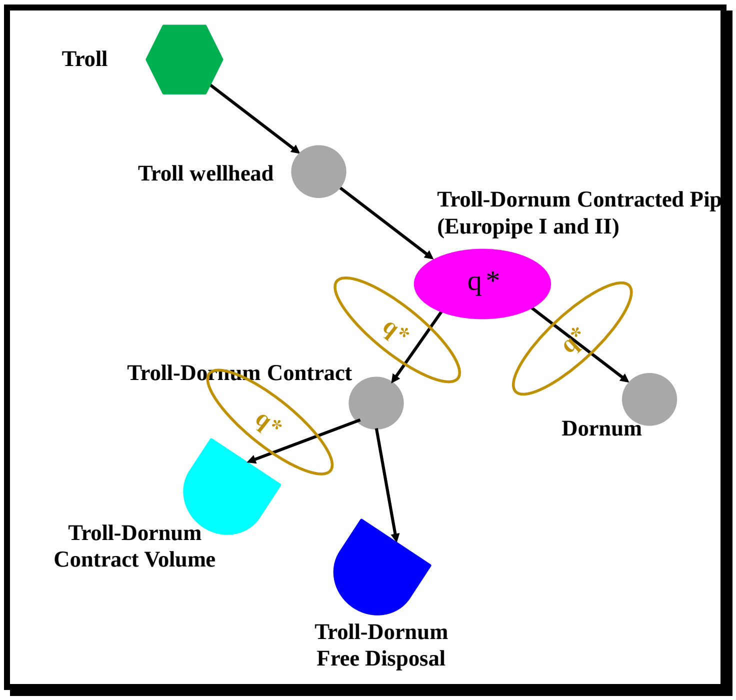


Using the free disposal node, we easily construct a model that the deliveries from Troll to Dornum will be at least as large as the contract quantity q*, and any additional deliveries would be spot deliveries on mutually attractive buy-sell terms between the producers, the pipe, and the customers in Dornum. That would motivate the network structure in Figure 49. This is a rather ingenious way to represent all the contracting phenomena at work from Troll to Dornum:

- Mandated take of at least the quantity q*.
- Calculation and makeup of the economic loss (if there is any) to force the contracted quantity q* into that market. If there is a better source than the contract source, there is an economic loss from the contract, and this construct will calculate it and offset it so that the contract is honored. This allows us to assess whether the contracts are imperiled in the market or whether they might persist.
- Voluntary take of volumes above the contract volume. Such voluntary take requires zero offset of loss. It is a spot transaction.
- Analysis of different (i.e., ramped up, ramped down) contract volumes to test their sensitivity.

- Disposal of overcontracted gas, if any, into the local spot market. Certain supply scenarios might have overcontracted gas, and certain might not.

Figure 49: Dornum with Free Disposal—Physical Gas ≥ Contract Gas

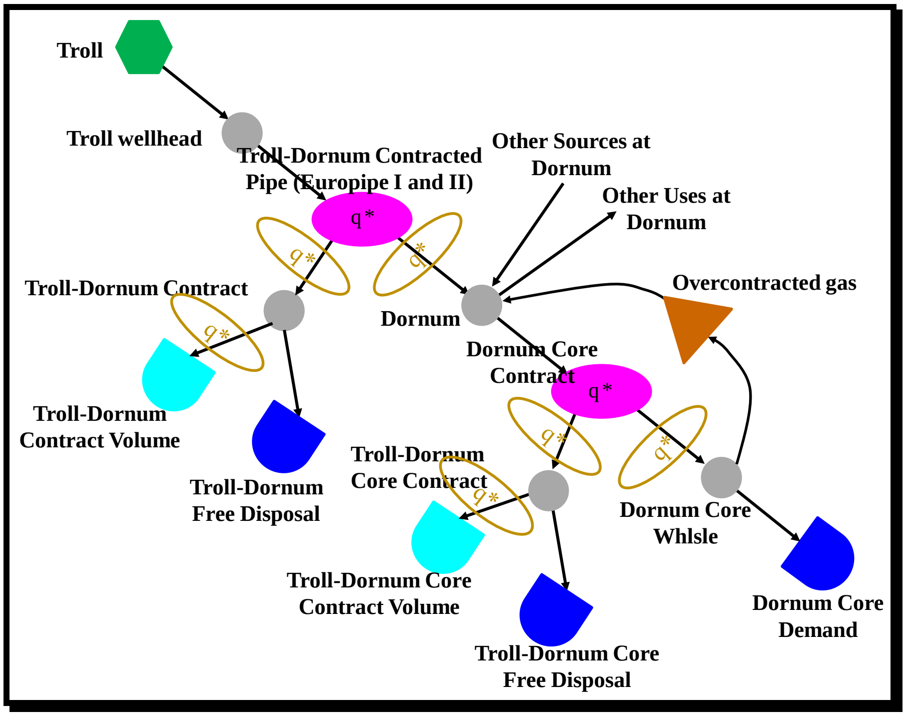


Finally, and this is **very** important, suppose that the take or pay contract with the Troll producers and the Europipe owners is backstopped by core customers in Dornum. That is, the local distribution company in Dornum has decided that it will warranty the take or pay contract on the backs of its core customers. This happens **all the time** in gas utilities. Any obligations utilities have to transporters or producers are warrantied on the backs of their core ratepayers who are captive to the utility. This type of construct actually requires that the contracted gas "flow through" the contracted volume q* at Dornum to the German core ratepayers in this case.

This type of construct—"flow through" take or pay contracting—ubiquitous in the real world in Europe, Africa, Russia, and Asia, is very easy to represent using network equilibrium concepts. See Figure 50 for a simple network representation of this complex and sophisticated "flow through" model of contract demand to core ratepayers. Notice that there is a contract volume q* from Troll to Dornum that is satisfied. The quantity q* is forced into the Dornum market from Europipe I and II as the contracts specify.

However, there is other gas coming in and going out of the Dornum market from other sources as well. The Dornum market is bigger than the contracted gas in that market. Such inflow and outflow clearly represents exchange and trading of that contract gas as well as spot gas at Dornum, while at the same time mandating downstream delivery of the contract gas to the core ratepayers at Dornum. This type of exchange and trading happens **all the time** in the real world, and it is very important to represent it as this construct does. Gas is traded and exchanged all the time, knowing that the combination of exchanged and contracted gas in a market must be sufficient to meet further downstream contract obligations. A model that lacks this trading and exchange feature is a doomed, problematic model.

Figure 50: What if the Contract Is with Core Customers



The contract with the core customers at Dornum draws from the Dornum citygate market and mandates that at least the quantity q* be delivered to core customers at Dornum. It de facto mandates that the core customers will draw at least the contracted quantity q* from the Dornum market. If there is any excess, it is transported back to the Dornum market where it is liquidated, as indicated by the orange triangle pointing from retail to wholesale in the Dornum market, ostensibly to another source or use, which is a spot or contract source.

Notice by virtue of this construct that the quantity q* is contracted at Troll and is delivered all the way through the system to Emden core ratepayers. It proceeds through the citygate market at Dornum and may be exchanged or traded as desired in that market. Then the quantity q* is subsequently sent downstream to the core Dornum customers.

Reflection on this type of network structure and node capability will convince one that this structure in fact truly represents take or pay contracting from any given source such as Troll to any given destination such as Dornum core ratepayers no matter how many intermediate wholesale markets those contracted volumes are obliged to flow through. This is an immensely powerful structure for representing and understanding the impacts of take or pay contracting in world gas pipeline and LNG transportation systems. We would seriously question whether methods other than network equilibrium methods could address the problem in this generality, convenience, and accuracy.

There is one other point to be made here, and that is the "price" of the take or pay contract. The notion of a mandated take at a given point in a market means that that gas must be taken no matter what the pricing terms in the contract. **It is the mandatory take that impacts and distorts the market.** The "price" terms lie outside the market. The contracting party is stuck with the price on the face of the contract. If the Dornum core customers agree to a take of q*, the pricing terms really do not matter in the markets per se. The price terms are "side payments" outside the market. Those customers are stuck with the Troll gas through the take or pay contract no matter what the price. Certainly gas can be exchanged out at points upstream and midstream in the system, but that Troll quantity will be delivered to Dornum core customers.

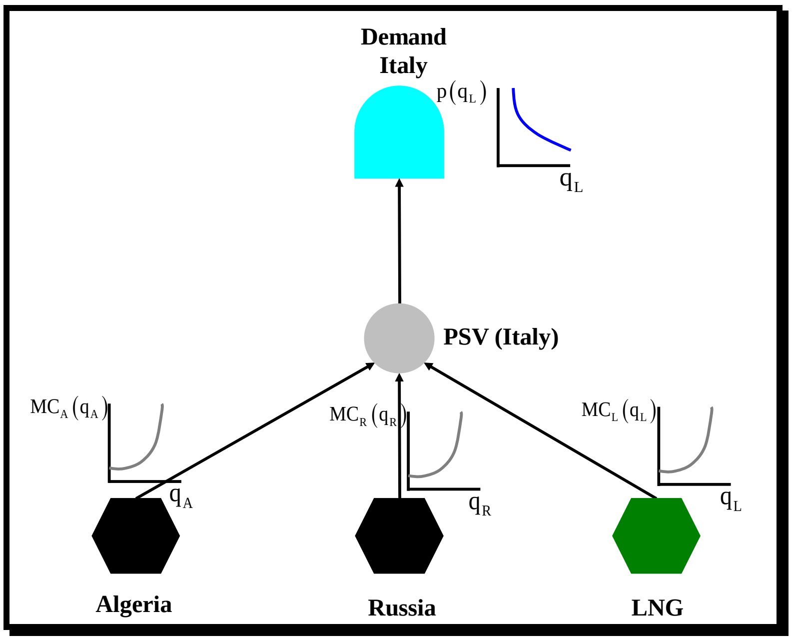
We have presented this protracted example here to attest to the power of network microeconomic modeling for otherwise difficult problems such as take or pay contracting. It is so transparent and obvious what is going on and why it is correct. It is also obvious that this type of structure would be difficult or impossible to represent in other methodologies.

10 A NOTE ON MARKET POWER

Market power (generally on the supply side but equally valid on the demand side) has received substantial attention in the past four decades. Complementarians have implicitly boasted that they can analyze market power (perhaps implying that the economic approach might have some difficulty or that the complementarity approach might have some measure of uniqueness). There are a couple of reasons such an argument would lack credibility. We are going to illustrate why using a simple example from our personal experience.

The largest electricity generation company in Italy has been a client of ours. During that assignment, we learned that Italy buys part of its gas from North Africa (primarily from Algeria) through the TransMed undersea pipeline from Algeria to Tunisia to Sicily to Italy and points North; part of its gas from Russia via pipelines from the Northeast via Uzhhorod at the Slovakia-Austria border; and part of its gas via LNG import onto the boot of Italy. The company reasoned that since Algeria and Russia had 90 plus percent of their market that they were in fact captive to market power and price setting by Algeria and Russia. They conjectured that their situation looked something like that in Figure 51. (The name of the Italian trading hub is PSV, which stands for "Punto de Scambio Virtuale." We use that to represent the Italian gas market.)

Figure 51: Potential Nash-Cournot Duopoly into Italy



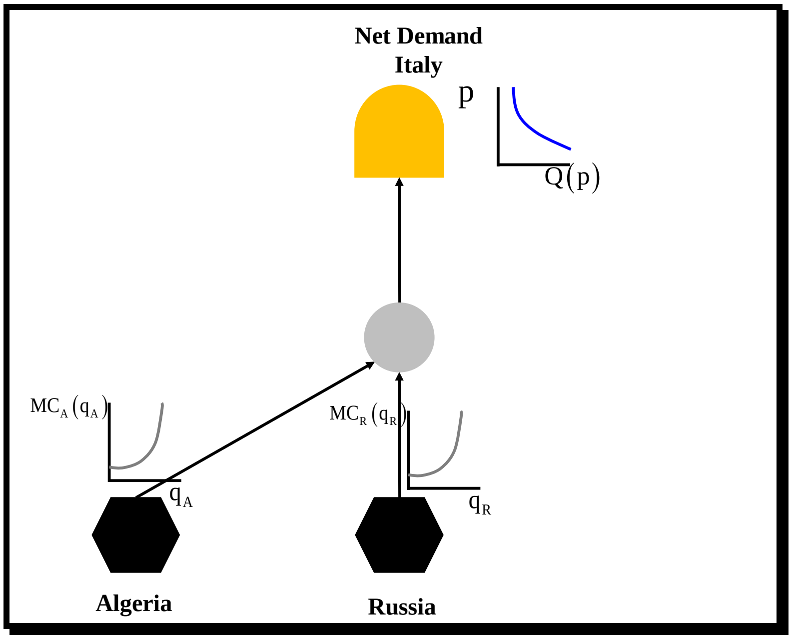
The theory of monopoly and oligopoly is lucidly clear (and quite heavily considered in the economic literature. James W. Friedman's book is definitive on the subject, and it has been around for 35 years). We appeal to the Nash-Cournot characterization of duopoly behavior. The Nash-Cournot theory envisions two large suppliers, collectively large enough to be a Stackelberg monopoly, selling product into a market, as represented by a demand curve. The two large suppliers (Russia and Algeria) each has a "marginal cost curve" for its production. The two large suppliers have such a large position in the market, each realizes that he can by his unilateral action restrict output and thereby increase the price in the market. Each has what economists call "market power." Each can, by his unilateral output decision, affect the price.

In the figure, there is a small supplier in the wings (called the competitive fringe) that competes with the postulated duopolists, namely landed LNG. The competitive fringe (LNG), because every player in it is so microscopically small, cannot by its unilateral action affect the market price. The fringe is a price taker. It is as if the fringe is comprised of an array of small producers, and each is individually too small to impact price by his unilateral actions. (This is called the "atomistic supplier" situation. Each supplier in the fringe is no bigger than an atom.)

The competitive fringe takes the price from the market as given and decides how much to produce based on his marginal cost curve. That is, he calculates how much he will supply S(p) at any given price p. The diagram also shows the demand, the "market," at the top. The market is characterized by a price dependent demand curve. The quantity consumed Q(p) is an explicit function of price.

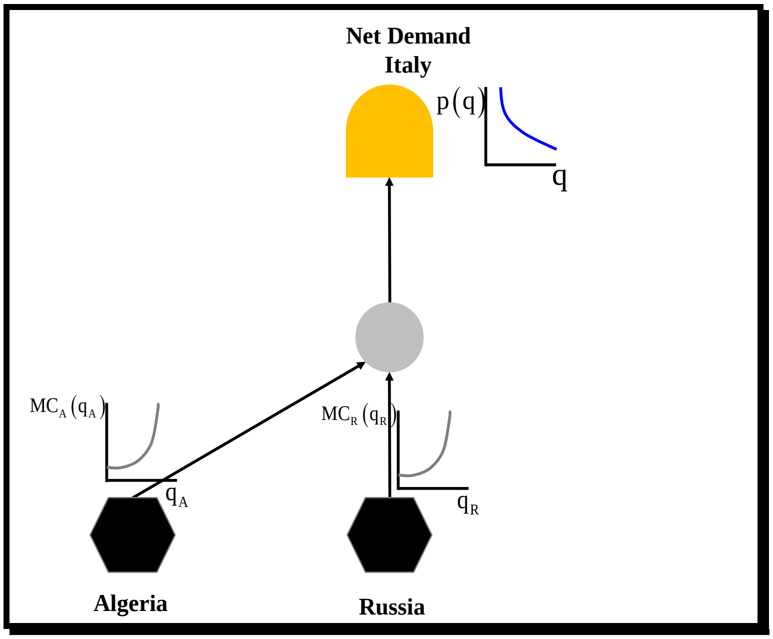
Because the consumer is a price taker, and the competitive fringe is a price taker, we can think of the net demand facing the duopolists. If for each level of price p, we start with demand at that price D(p) and subtract competitive fringe supply at that price S(p), we can quickly construct what is termed the "net demand to the duopolists," i.e., Q(p) = D(p) - S(p). Under this construct, we reduce the diagram in Figure 51 to the form in Figure 52.

Figure 52: Net Demand Facing the Duopolists



Keep in mind, the demand curve in this market is gross demand (gross demand in Italy) net of supply by the competitive fringe (LNG). That is important. The nature of this net demand depends on the competitive fringe as well as the market. It helps with subsequent discussion to invert the demand curve, expressing p as a function of q, and represent the duopolists' problem as in Figure 53. Notice that there is a large, duopolistic supplier selling into a net demand function, a demand function that represents gross demand **minus** competitive fringe supply.

Figure 53: Indirect Net Demand Facing the Duopolists



What is it that these duopolists will be expected to do? What stops them from benefiting from their market power? Assuming that each duopolist strives to maximize his profit, i.e., maximize the amount of money he makes, what would we expect him to do? Clearly each would set his level of output q so that it maximizes his individual profit:

$$\text{Algeria:} \quad \underset{q_A \geq 0}{\text{MAX}} \quad q_A p(q_A + q_R) - TC_A(q_A) = q_A p(q_A + q_R) - \int_0^{q_A} MC_A(x)\,dx$$

$$\text{Russia:} \quad \underset{q_R \geq 0}{\text{MAX}} \quad q_R p(q_A + q_R) - TC_R(q_R) = q_R p(q_A + q_R) - \int_0^{q_R} MC_R(x)\,dx$$

Notice how tightly the theory is bound to the notion of a net demand function. The economic "answer" to the foregoing duopolists' problems have been very well known for many, many years, and it is instructive to write it mathematically. First, we transform to Luenberger form:

$$\text{Algeria:} \quad \underset{q_A \geq 0}{\text{MIN}} \quad -q_A p(q_A + q_R) + \int_0^{q_A} MC_A(x)\,dx$$

$$\text{Russia:} \quad \underset{q_R \geq 0}{\text{MIN}} \quad -q_R p(q_A + q_R) + \int_0^{q_R} MC_R(x)\,dx$$

The two Lagrangians are:

$$L(q_A) = -q_A p(q_A + q_R) + \int_0^{q_A} MC_A(x)dx - \mu_A q_A$$
$$L(q_R) = -q_R p(q_A + q_R) + \int_0^{q_R} MC_R(x)dx - \mu_R q_R$$

The Kuhn-Tucker conditions for each of the two problems are:

$$\frac{\partial L(q_A)}{\partial q_A} = -p(q_A + q_R) - q_A p'(q_A + q_R) + MC_A(q_A) - \mu_A = 0$$
$$\mu_A q_A = 0 \qquad \mu_A \geq 0 \qquad q_A \geq 0$$

$$\frac{\partial L(q_R)}{\partial q_R} = -p(q_A + q_R) - q_R p'(q_A + q_R) + MC_R(q_R) - \mu_R = 0$$
$$\mu_R q_R = 0 \qquad \mu_R \geq 0 \qquad q_R \geq 0$$

As before, what if you had an analytic functional form for both of the marginal cost curves and for the demand curve? Couldn't you substitute that functional form into the Kuhn-Tucker conditions and solve them analytically? Wouldn't that be preferable to defaulting to setting up and solving complementarity equivalents of Kuhn-Tucker conditions? Of course it would. Just as in Section 7, it is preferable to make the maximum effort to solve these Kuhn-Tucker equations analytically before moving to a numerical algorithm, particularly a brute force solution algorithm such as a GAMS or AMPL complementarity algorithm. We will not pursue this theme further here because the coup de grace for simple Nash-Cournot solutions emanates from another source.

The complementarity equations for the foregoing two optimization problems can be written:

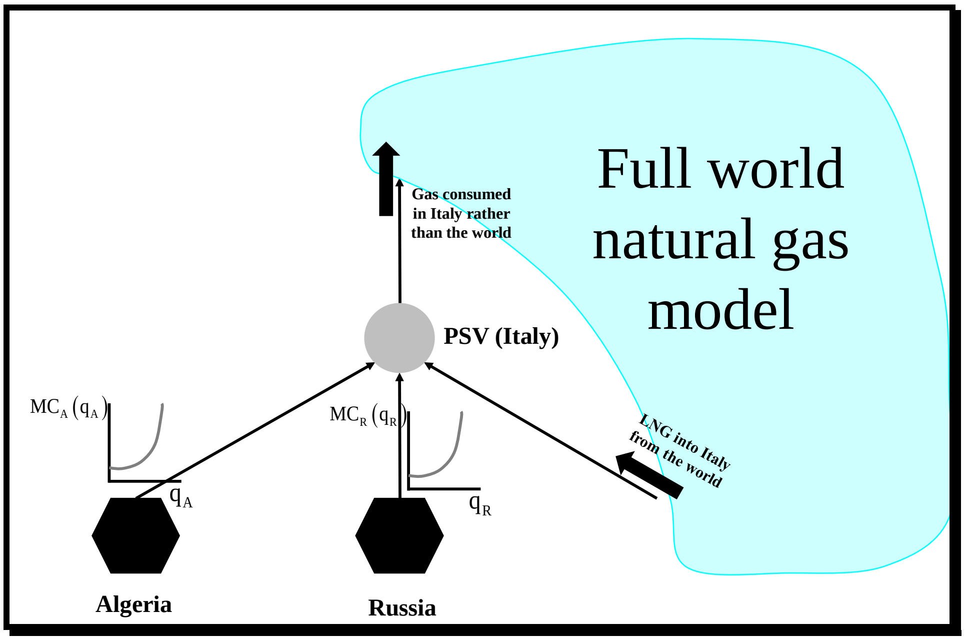
$$\left[-p(q_A + q_R) - q_A p'(q_A + q_R) + MC_A(q_A)\right] q_A = 0$$
$$-p(q_A + q_R) - q_A p'(q_A + q_R) + MC_A(q_A) \geq 0 \qquad q_A \geq 0$$

$$\left[-p(q_A + q_R) - q_R p'(q_A + q_R) + MC_R(q_R)\right] q_R = 0$$
$$-p(q_A + q_R) - q_R p'(q_A + q_R) + MC_R(q_R) \geq 0 \qquad q_R \geq 0$$

Sure, they are indeed complementarity equations. Sure, one could program them, bundle them, "send them to a solver," and access the answer. That is not going to be insightful or useful for a reason that will be made clear shortly. Complementarity isn't the primary difficulty here. The difficulty is the naiveté of the formulation. When one removes the naiveté of the formulation, we are right back to substantial size network problems, and complementarity is not the way to solve them.

What is the true and correct situation? The true and correct situation is **not** that in Figure 51. Not even close! First of all, there is no way to observe a competitive fringe supply function into Italy that represents LNG. Such a function does not exist and cannot be written or estimated as required by the simple Nash-Cournot duopoly model. Quite the contrary, the supply function for LNG into Italy is the output of a very large, very detailed, very sophisticated full world gas trade model, i.e., the structure in Figure 54. These ostensible Nash-Cournot players are embedded in a large, multi-thousand node, sophisticated supply, transportation, and demand model as illustrated in Figure 54. The competitive fringe is the solution to that full world model.

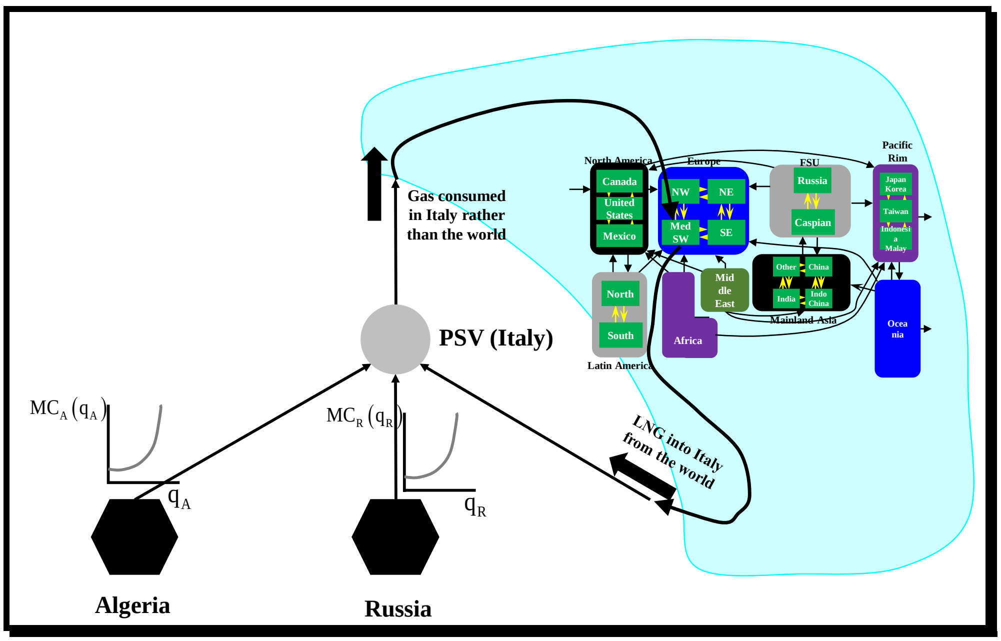
Figure 54: Potential Nash-Cournot Duopoly into Italy Requires a Full World Gas Model



Obviously, three of the interconnected regions of this world supply-transportation-demand model will have to be Algeria, Russia, and Italy. The LNG into Italy and the demand for gas in Italy are points of entry and exit within this large, full, interconnected world natural gas model. We attempt to illustrate this detailed regionality using Figure 55. We depict schematically (not exactly) where the Italian LNG imports will emanate, and we depict schematically (not exactly) where the demand in Italy will occur.

The two important insights are this: (1) there is **<u>no</u>** supply function for imports into Italy and (2) there is **<u>no</u>** demand function for consumption in Italy plus export to points beyond. There is a monster world gas model in which consumption in Italy competes with exports from Northern Italy to Switzerland, Austria, or Germany. The demand curve at PSV is not only the consumption in Italy but, in fact, all the supply, transportation, and consumption downstream from Italy. The supply curve at PSV for LNG is not only the supply into Italy but in fact all the supply, transportation, and demand upstream from Italy. It is a heck of a lot more complicated than positing a simple scalar indirect demand function as Nash-Cournot models are wont to do! In fact, you need a full world gas model in order to have the slightest hope of measuring or quantifying market power, whether monopoly, oligopoly, monopsony, or oligopsony. The approach of centering the analysis on a contrived demand function p(q) and its derivative p'(q) for something as sophisticated as European gas supply and demand or even worse OPEC net demand is felicitous. These overly simplified models of market power carry no real insight. Their insights are limited by the context of their assumed supply and demand curves, which do not exist as such. They are not supply and demand curves. They are full models of the world and their image at the injection and withdrawal points within Italy. EIA needs a fully operational competitive model of the world, including a large transportation matrix, in order to even begin to approach the issue of market power, duopoly, and the like. It does not matter if complementarity or network microeconomic people can solve trivial pencil and paper problems. It matters that network people can solve very large problems and deliver the competitive fringe supply and the demand information to the Nash-Cournot players. Only by doing this will the Nash-Cournot solution have any predictive or interpretive value.

Figure 55: Potential Nash-Cournot Duopoly into Italy Must Be Part of Structured World Gas Model



Complementarity is clearly no silver bullet for market power, not at all. It faces exactly the same considerations as microeconomics. In fact, microeconomics is better, able to calculate whether a player or players might have market power by testing how the world will respond to production decisions by those players. This is far preferable to arbitrarily assuming a local demand function and a local competitive supply function.

11 AN EMBEDDED, ENDOGENOUS MODEL OF COMMODITY STORAGE

Storage of gas, refined products, water for hydroelectric, or electricity in batteries/capacitors is an intrinsic player in the marketplace. (In this section, we use gas storage as our discussion example. Most or all of the concepts extrapolate to oil, water, and electricity.) We know that storage adds to gas supply during time of peak (more gas available to deliver to time of peak demand) and adds to gas demand at time of off-peak (buys gas at time of off-peak to service the forthcoming peak need). As such, storage puts:

- Upward pressure on price during time of off-peak. By increasing gas demand at time of off-peak, storage elevates off-peak price.

- Downward pressure on price during time of peak. By increasing gas supply at time of peak, storage depresses on-peak price.

Keep these pricing ideas in mind, for they are central to understanding storage.

There are a number of applications of storage in energy markets, and they impact markets and markets impact them:

- Natural gas storage for seasonal variation
- Natural gas storage for short term variation
- Petroleum product storage for seasonal variation
- Petroleum storage for strategic reasons
- Reservoir hydro storage for seasonal variation
- Pumped hydro storage for short term variation
- Batteries and capacitors for very short term variation, a growing area

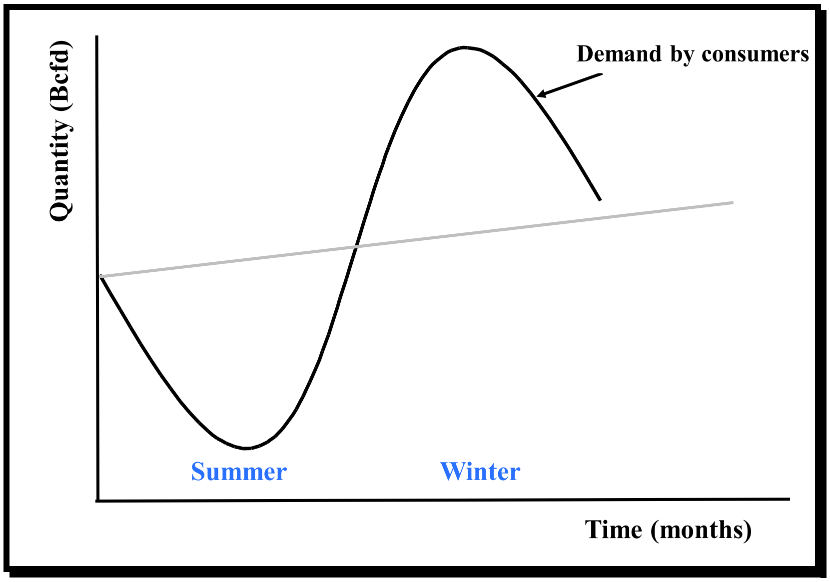
Modeling systematic commodity storage (e.g., natural gas and liquid storage) is actually more difficult than modeling phenomena such as reservoir hydro storage because, in the gas and liquids case, one has to make two decisions—injection and withdrawal. In the case of reservoir hydro, there is only one decision to make—when to generate. If the storage model can represent the natural gas and liquids storage problem, the hydro reservoir problem (pumped as well as seasonal) is a simple special case in which injection is eliminated and water materializes via hydrology.

We are going to explain the gas and liquids storage module EIA needs here. Specialization to pumped or reservoir hydro and other storage mechanisms is straightforward.

11.1 Storage Reduces Seasonal Swings

The demand (consumption) of a commodity in many cases follows a seasonal pattern. Natural gas is an example. Natural gas consumption through time occurs as shown in Figure 56. We note that consumption in the summer is low because there are no homeowners or establishments that require heating in the summer. Similarly, consumption in the winter is high because the weather is cold, and there are many homes or establishments that must be heated for safety and convenience. We also note in the figure that there is generally an upward trend in consumption forward in time, as indicated by the obliquely tilting line in Figure 56.[36]

Figure 56: Natural Gas Consumption Is Seasonal



If a producer had to meet this real-time pattern of demand, several things would have to occur. First, he would have to size his total production capacity to the absolute peak. Secondly, he would have to shut down much or all of his production capacity during the summer (off-peak). There would be no one to accept and consume his product. He would have to run at higher and higher capacity during the winter peak period, running at full capacity only on the peak winter day. Figure 57 illustrates the situation without storage.

What happens when storage enters the picture? During the summer, people are able to draw from real-time gas deliveries to inject gas into a storage facility **plus** serve the low level of consumption that occurs during the summer. In effect, the market buys gas to inject into storage plus serve the seasonally low summer demand. During the winter, people are able to withdraw gas from the storage facility to supplement real-time deliveries so as to serve the peak winter market in total. Withdrawal from storage plus real-time deliveries are sufficient to satisfy the peak demand. The situation becomes that in Figure 58.

[36] For purposes of this section, we are not considering the stochastic variation in demand, which is important for storage, but only the systematic, predictable, known seasonal variation in demand.

Figure 57: The System Must Be Sized to the Peak

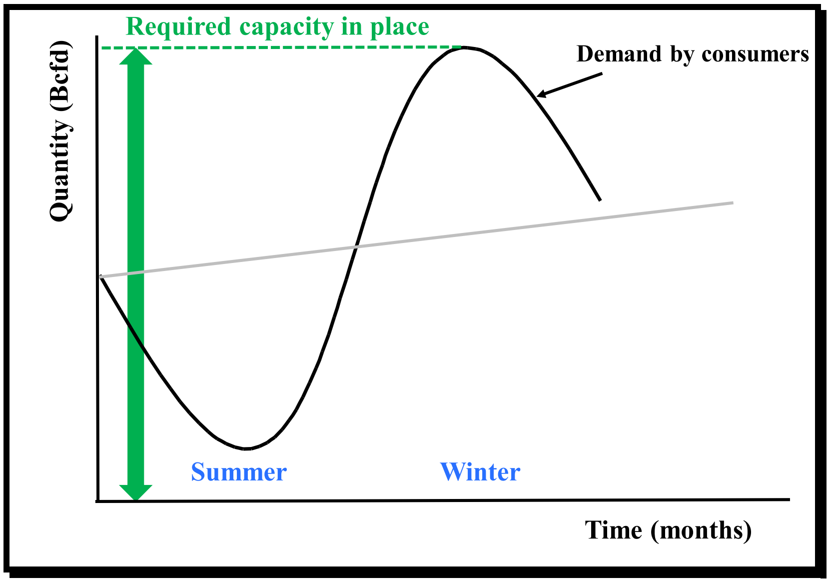
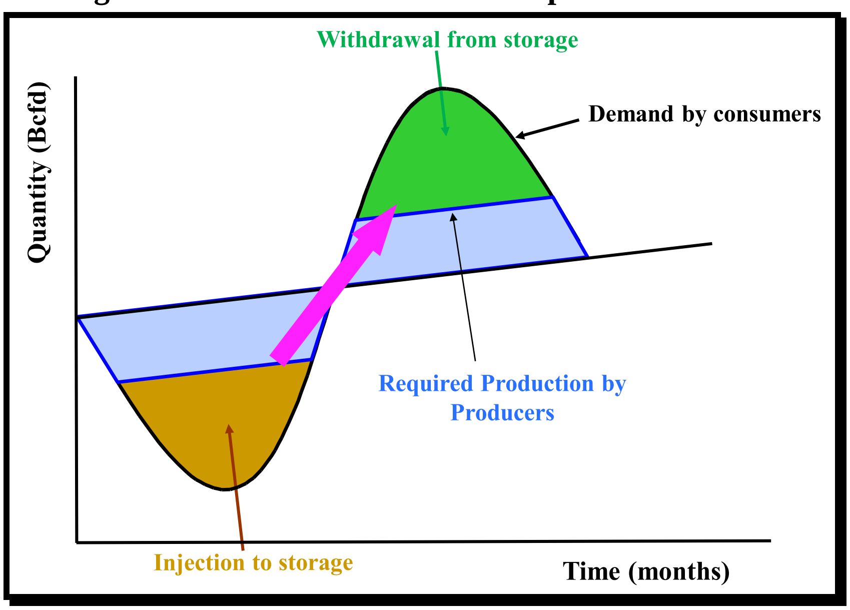


Figure 58: Natural Gas Consumption Is Seasonal

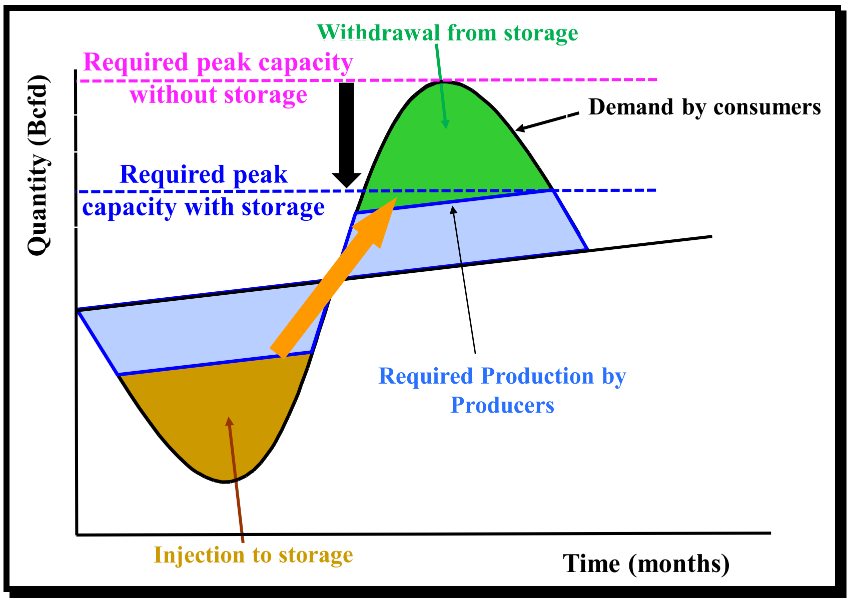


We note in Figure 58 the highly varying demand by consumers, the outside black curve in the diagram. The demand by consumers is the same as it was in Figure 56. We also see that during the summer at the left, there are deliveries to customers **plus** injections to storage. The quantity of gas that is provided by producers to the market during the summer at the left is represented by the solid blue line. It is comprised as shown of the actual consumption by consumers during the summer plus the injections to storage during the summer, which is represented by the brown area at the left during the summer. Summer consumption is **higher** than the very low consumption by customers, the difference during the off-peak period being delivered for injection into the storage facility.

We note in Figure 58 during the winter at the right that the consumption is represented by the very high black line at the top. However, the delivery by producers is the solid blue line above the annual average upward trending line but well below the black line. It is, in fact, the blue line at the right. That real-time delivery by producers indicated by the blue line is augmented by withdrawals from storage represented by the green area during the winter. Deliveries through the pipeline (the blue line) are augmented by deliveries from storage inventory (the green area) so that the sum total satisfies demand by consumers. The notion here is that storage buys gas at time of off-peak and sells gas at time of peak, decreasing the need for real-time deliveries at time of peak but increasing the need for real-time deliveries at time of off-peak. Storage "levels the load" that producers have to meet. The levelized load that occurs as a result of storage, the load that producers have to meet, is the blue line in Figure 58. The effect of storage is that producers meet a much less variable, must less volatile load than they would without it.

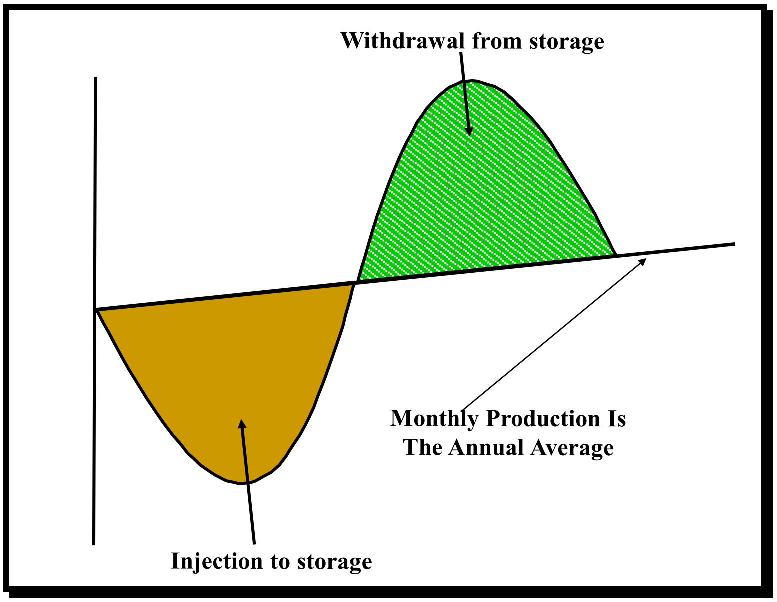
The impact of storage on the magnitude of producer capacity in place is indicated in Figure 59. The peak that producers must maintain is far lower in the "with storage" case than the "without storage" case. This can dramatically affect real-time prices as well as annual average prices and costs. It dramatically affects the capacity factor on the various supply and transportation facilities. This is why storage is potentially so important.

Figure 59: Natural Gas Storage Reduces Capacity in Place



It is well to consider that if there were an oversized storage system, one could completely buffer the demand variation. One could in the summer store all the gas one needed in the winter such that the quantity delivered by producers in the summer would be equal to the quantity delivered by producers in the winter. In such a situation, represented by Figure 60, the summer-winter price differential would collapse to pure variable cost plus holding cost, the entirety of peak deliverability having been buffered by storage during the summer. This is meant to illustrate that the size of the storage reservoir and the rate at which gas can be injected to and withdrawn from the storage reservoir are very important. We shall see that in the ArrowHead storage model.

Figure 60: Natural Gas Consumption Is Seasonal



11.2 Storage and Wellhead Deliverability

What does storage mean for crucial issues such as resource depletion or cumulative reserve additions? Without storage in place, one would have to add sufficient short term wellhead deliverability in order to meet the peak consumption. Wellhead deliverability added is represented by the top lines in Figure 61 for the second winter peak. By contrast, if there were sufficient storage to inject significant quantities during the summer and withdraw them during the winter, the production requirements met by the producer would be the black, smooth, gradually increasing line in the middle. The wells would run at full capacity, and the market would only need enough wells to hit the black lines in the second period. You would need far less peak deliverability at the wellhead. You would run your wells close to 100 percent capacity factor. The key observation is that the market would need far fewer producing wells. Resource base depletion would be retarded. Wellhead costs would be lower. Figure 61 illustrates the effect your storage model must have. It must carefully quantify and measure the reduction in peak wellhead deliverability that would occur by virtue of storage being present in the market. The model outlined here does precisely that and thereby gives more accurate predictions of resource needs and depletion.

As we infer from Figure 61, storage decreases the high value (and high marginal cost) of short term deliverability swings. Storage delays the need to drill and prove new reserves. This is a marketwide impact of storage, and the storage model we have built and recommend to EIA takes this fully into account when properly embedded in a regional or world model.

When we observe the operation and capacity, we see the situation for the supply and transportation system as being that in Figure 62.

Figure 61: Storage Reduced Cumulative Reserve Additions

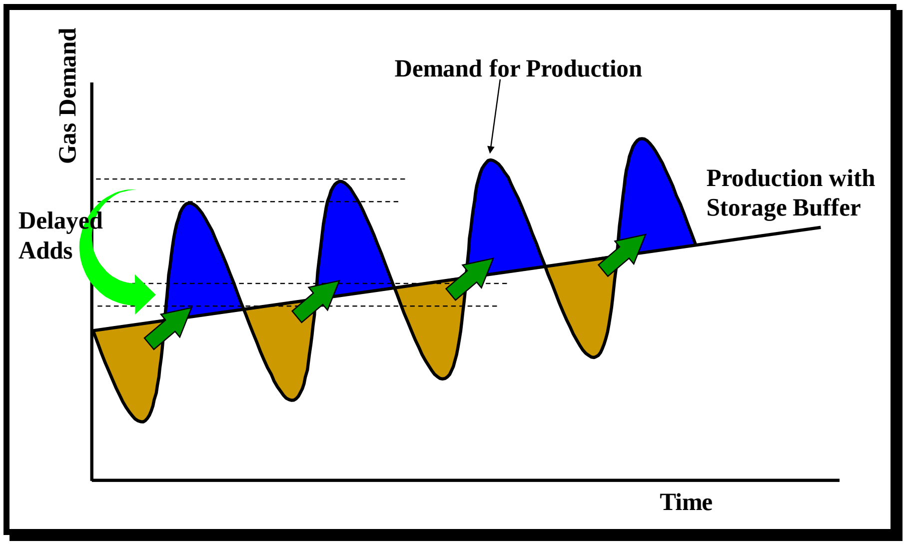
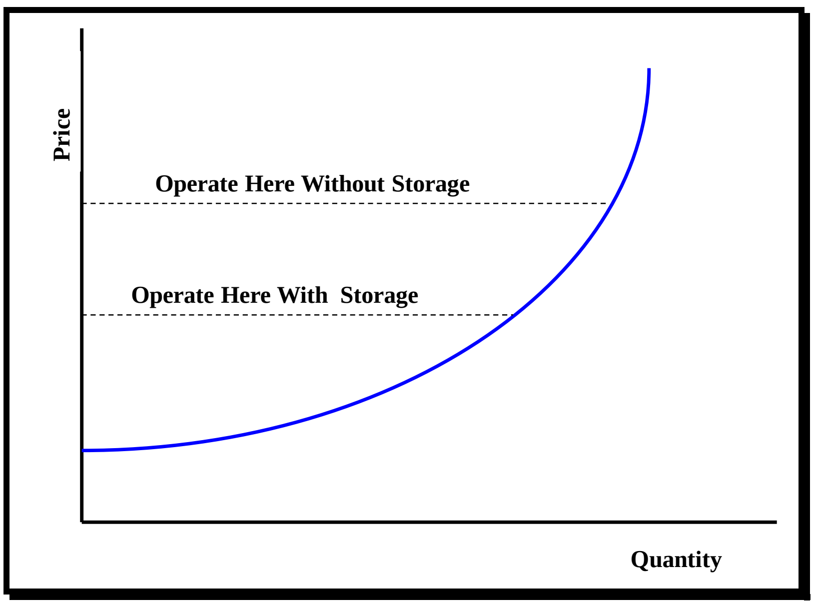


Figure 62: Storage Delays Resource Depletion and Downsizes Transportation Infrastructure

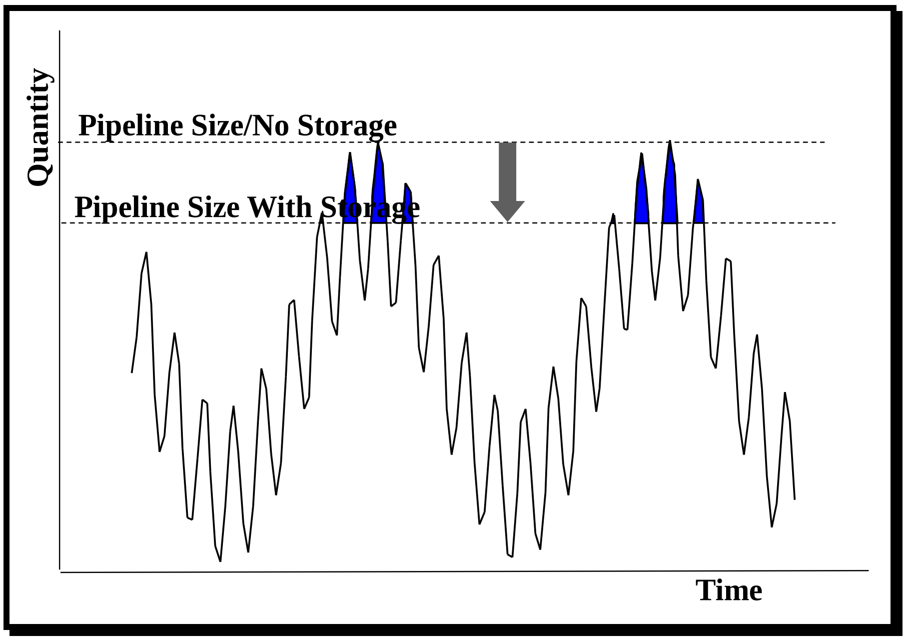


As we emphasize in the diagram, storage retards rate of exploration and production (E&P). Producers do not have to explore just to create peak deliverability. Storage allows a high capacity factor, just in time production pattern. The storage model embedded in a full network model represents the interplay of storage and the need for incremental peak deliverability.

11.3 Storage and Transportation/Transmission Deliverability

The situation is very much the same for transportation as it was for resource production. It is obvious from the discussion surrounding Figure 61 that market area storage contiguous to customers will reduce the quantity of inbound pipeline capacity that is required. The ArrowHead storage model represents that phenomenon. However, transmission sizing also interacts with monthly and shorter-term demand variation. Figure 63 illustrates how short-term buffering of short-term demand swings allows the inbound pipe to be undersized relative to the time of peak demand, the short term swings buffering in and out of market area storage. We see in this diagram the longer cycle demand variation and superposed on that the shorter cycle demand variation. Our storage model represents this phenomenon as well, buffering inbound pipeline capacity over short-term demand variations and thereby downsizing inbound pipeline capacity requirements. The recommend ArrowHead storage model represents this phenomenon as well. This is also true of LNG tankers who, once on the open ocean, can steer for the high gas price areas or steer for the areas with storage which can accept their cargoes any time.

Figure 63: Transmission Sizing Interacts with Monthly (and Shorter-Term) Storage

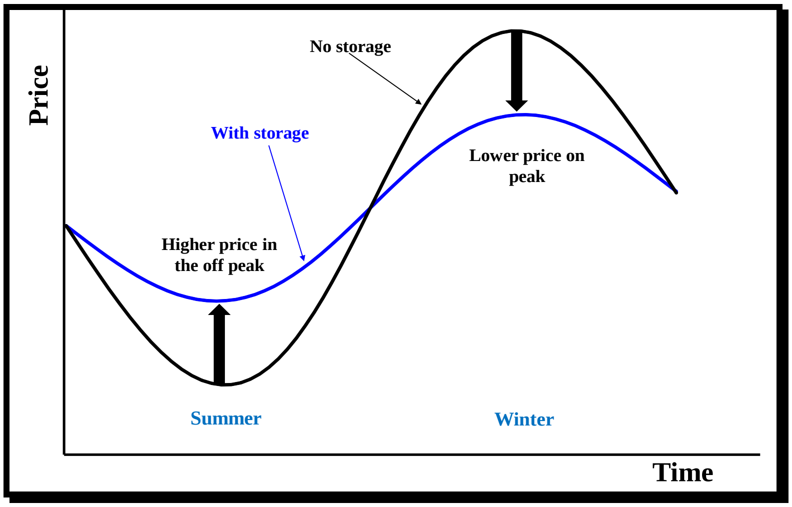


Storage clearly allows the market to undersize inbound transmission (relative to what it would be without storage). Storage is analogous to a “peaker” in electric generation. Storage saves **lots** of money and dramatically downsizes the necessary system, and it alters market prices and quantities.

11.4 Storage Reduces the Peak-Off-peak Price Differential

We have discussed, but it is worth strongly reiterating, that storage reduces the differential between the price that occurs at time of peak and the price that occurs at time of off-peak. Figure 64 illustrates. The storage model must lead systematically to this type of result, and the one we articulate here does. (Storage models that extrapolate history often do not. See below.) Furthermore, if there is a surfeit of storage, the peak-off-peak price differential will fall to the variable cost of storage, including inventory holding cost. The model we articulate here will do that. Lesser amounts of storage in place (the typical rule around the world) will impact but not collapse the summer-winter price differential and will retard the quantity of wellhead peak deliverability and pipeline peak deliverability required.

Figure 64: Storage Reduces the Peak-Off-peak Price Differential



11.5 The ArrowHead Storage Model

This section summarizes the correct and proper way to represent a storage asset in an economic context. The fundamental concept underlying the proper model of storage is that every storage asset resides at a hub. The storage asset buys gas from the hub at the prevailing market price, and it sells gas to the hub at the prevailing market price. In the network microeconomic context, there is a storage asset at the market hub contiguous to that storage asset, and it buys and injects gas from the hub, places it into inventory, and sells gas from that inventory back to the market at certain other times. The storage asset owner will strive to buy and sell gas so as to maximize his profit from so doing.

There are storage assets located in a distributed fashion scattered throughout the entire world gas market, each one situated as in Figure 65.[37] Some storage lies in the upstream, some lies along pipes, and some lies in market areas. All are situated with respect to local market hubs as shown in Figure 65.

11.6 Mathematical Summary of the Storage Module

To illustrate how the EIA storage model must work, consider a highly simplified situation in which there are N time intervals of equal length. For the sake of discussion, let us assume that they are monthly. Figure 66 illustrates the time structure for this summary discussion of the storage model.

[37] This storage model has been reviewed and given a clean bill of health by Prof. Jeffrey Williams (Stanford and UCD), the preeminent storage economist in the world in January 2005 in the ArrowHead Padd 5 Oil and Refined Products Model and the ArrowHead North American Gas Model.

Figure 65: The ArrowHead Storage Model

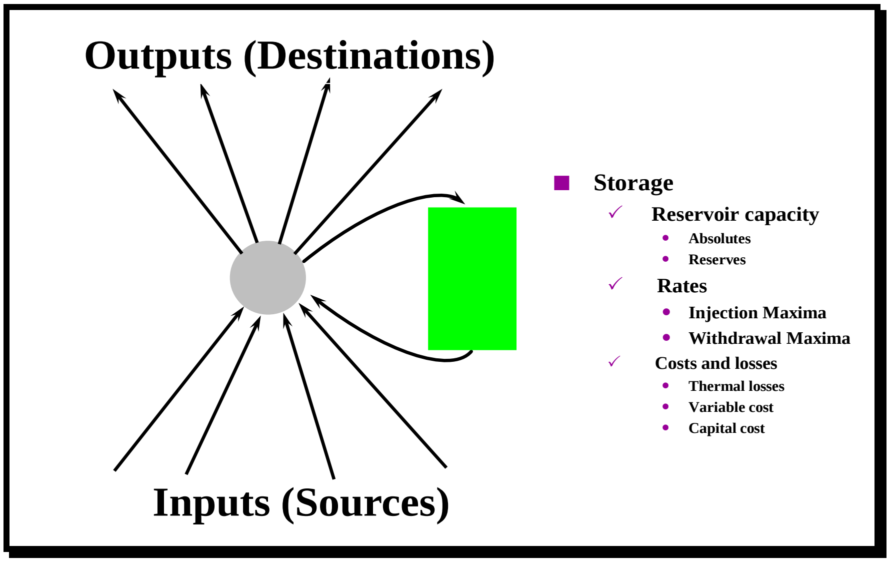
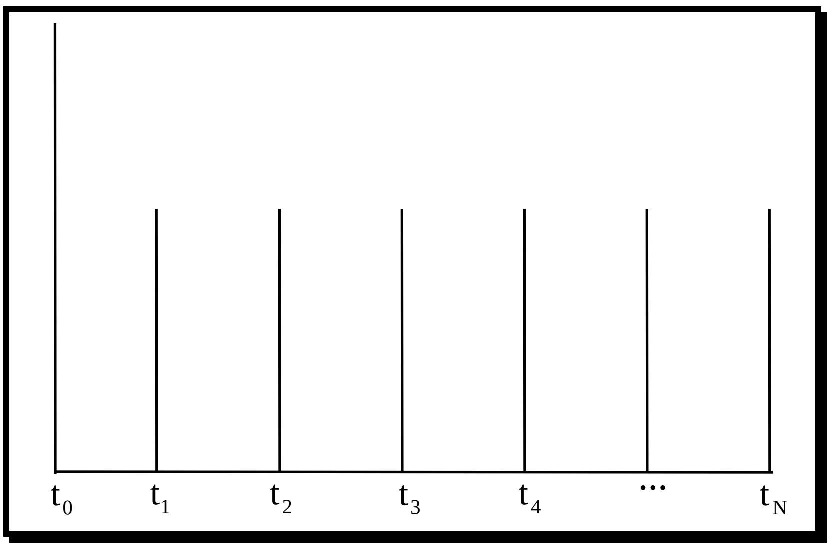


Figure 66: Time Points of Equal Length for Storage Model Illustration



Even though the time intervals are months in this discussion, they could just as well be weeks or days for gas and hours or portions of hours for electricity. We will require a bit of mathematical notation to summarize the storage model:

N=number of time points under consideration.
$I(t_0)$=inventory in the storage tank at the beginning of the model horizon.
I(t)=inventory in the storage tank at time t, $t=t_0, t_1, t_2, \ldots, t_N$.
ε_S=efficiency of injection (1 minus losses during injection.)
s(t)=quantity purchased from the local hub and injected to storage and at time t=1,2,…,N.
ε_W=efficiency of withdrawal (1 minus losses during withdrawal.)
w(t)=quantity withdrawn from storage and delivered to the local hub at time t=1,2,…,N.

For a given injection and withdrawal strategy s(t) and w(t), it is straightforward to calculate the inventory that will be in the tank at every time point in the model horizon. That inventory is, for time point t_n, n=1, …, N:

$$I(t_n) = I(t_0) + \sum_{k=1}^{n} \left[\varepsilon_S(t_k) s(t_k) - w(t_k) / \varepsilon_W(t_k) \right]$$

The inventory in the tank at any time point t_n is the inventory present in the storage reservoir at time 0 plus cumulative monthly injections (adjusted for losses) less cumulative annual withdrawals (adjusted for losses) up until time t_n.

There are two concepts related to the size and capacity of the storage assets needed for our model:

R = maximum size of the reservoir.

ρ(t)=lower bound on working gas availability at time t (reserve requirement mandated by the regulator or reserves carried by the operator as a matter of practice). Regulators do not allow inventories to go to zero, and operators generally keep some level of reserves in their tank so that they can meet peak deliverability requirements.

The inventory in the tank at time t_i must be larger than the reserve requirement but smaller than the maximum size of the reservoir:

$$\rho(t_i) \le I(t_0) + \sum_{k=1}^{i} \left[\varepsilon_S(t_k) s(t_k) - w(t_k) / \varepsilon_W(t_k) \right] \le R, \quad i = t_1, t_2, ..., t_N$$

This will be the **<u>first constraint equation</u>** (de facto the production function) that will govern profit maximizing behavior on the part of the storage facility owner.

In reality, there is a maximum rate of injection (governed by phenomena such as injection horsepower on compressors) and a maximum rate of withdrawal (governed by the size of the outlet and the ability to deliver it into a pipeline system). The storage model needs to take account of these injection and withdrawal maxima. The simplest possible way to do so is to posit the constraints:

σ(t)=upper bound on injection rate at time t (simplified for illustration.)
η(t)=upper bound on withdrawal rate at time t (simplified for illustration.)

This implies the following constraints that govern profit maximization:

$$0 \le s(t_i) \le \sigma(t_i)$$
$$0 \le w(t_i) \le \eta(t_i)$$

This will be the **second constraint equation** (de facto the production function) that will govern profit maximizing behavior on the part of the storage facility owner.

The foregoing two phenomena govern the inventory level and how rapidly you are able to augment it or draw it down. We now must turn to the economic and behavioral part of the storage model. To do so, we need a few more mathematical terms:

ϕ_S = variable cost (handling cost) injection, excluding any inventory storage charges.
ϕ_W = variable cost (handling cost) of withdrawal, excluding any inventory storage charges.
D(t)=discount factor from time t back to time 0.

We are now prepared to write the equation for the profitability that the owner of the storage facility can achieve by selling and buying gas to and from the hub to which it is attached. The storage field operator takes the price of gas at the contiguous hub to which he is connected as given and maximizes profit by buying gas from the hub and selling gas to the hub. In particular, the storage field operator defines his profit according to the equation:

$$\mathrm{Profit} = \sum_{k=1}^{N}\left[p(t_k)-\phi_W\right]w(t_k)D(t_k) - \sum_{k=1}^{N}\left[p(t_k)+\phi_S\right]s(t_k)D(t_k) + \mathrm{Salvage}$$

The first summation is the discounted present value of profit the storage facility operator receives from sale of gas back to the local hub. Notice that it contains the price of gas minus the variable cost of handling. Losses upon injection and withdrawal are taken account in the constraint equations as we maximize this measure of profit. The second summation is the discounted present value of cost the storage facility operator bears from having to purchase gas from the local hub. Notice it contains the price of the gas plus the variable cost of handling. The third term is the salvage value of the inventory in the tank at the end of the model horizon.

The storage facility owner will then set his injection and withdrawal strategy over time so as to maximize profit subject to the two constraints developed above. That is, the storage field operator will set his injection and withdrawal schedule so as to solve the following profit maximization problem:

$$\mathrm{MAX} \quad \sum_{k=1}^{N}\left[p(t_k)-\phi_W\right]w(t_k)D(t_k) - \sum_{k=1}^{N}\left[p(t_k)+\phi_S\right]s(t_k)D(t_k) + \mathrm{Salvage}$$

SUBJECT TO

$$\rho(t_i) \le I(t_0) + \sum_{k=1}^{n}\left[\varepsilon_S(t_k)s(t_k) - w(t_k)/\varepsilon_W(t_k)\right] \le R$$

$$0 \le s(t_i) \le \sigma(t_i)$$

$$0 \le w(t_i) \le \eta(t_i)$$

This is the fundamental behavioral relationship governing storage operation no matter where or what time frame—profit maximization subject to injection and withdrawal maxima and subject to reservoir capacity and minimum reserve requirement. The losses, reservoir size, initial model inventory, and all other phenomena are explicitly considered and endogenized herein. One can see all those variables contained in the profit maximization problem. This model of storage, precisely the one EIA needs, allows storage to be sited in the model in disparate locations throughout the world—wellheads or production areas, intermediate areas along pipelines, and market areas contiguous to consumption regions lying downstream from pipelines. You drag and drop a storage node everywhere in the world that a storage asset exists, or you think could exist, and link it up to the local hub. It will operate as a profit-maximizing producer would, and it will impact volumes and prices of logistical assets as well as production and consumption assets in precisely the way summarized in the first part of this section.

With this construct, all the myriad storage operators everywhere in the world are explicitly assumed to be "players," i.e., agents. They all interact, and they all individually modify supply and demand patterns as indicated above, and in precisely the right way. Storage is intrinsically an "optimization problem within an economic problem," and EIA needs to represent it this way.

The ArrowHead storage model represents every storage asset in the world:

- Individually
- Independently

This is precisely as is true in the real world. By so doing, every storage node in the world thereby competes with and/or complements every other storage node and pipeline and supply source in the world. This is the quintessence of logistical modeling. Every storage node competes against every other node in the model to make money and retemporalize flows. Every storage node is a proactive, price taking, profit maximizing competitor. This is what renders network microeconomic modeling unique and accurate.

11.7 Summary

The ArrowHead endogenous storage model will buy and sell from the hub to maximize profit. Such behavior will serve to decrease price at time of peak and increase price at time of off-peak. That is the quintessence of storage. Furthermore, the proper endogenous operation of storage naturally decreases the incentive to build further storage. The model will find the right amount of storage, the amount that is profitable in a given region but no more.

One important observation: with this storage model, the storage injection and withdrawal schedule is not exogenous, it is endogenous. We have seen all too many times linear programming practitioners making the blithe assumption that storage asset operators will dispatch storage in the future the same way they have dispatched in the past. This is laughably wrong, and the recommended storage model will not do that. (We have seen such models turn summer-winter prices upside down in practice, the winter price being lower than the summer price, because of some scenario assumption like an abnormally warm winter or a hot summer. Such results cannot possibly be right.) The recommended model asserts that the storage asset owner will coldly and rationally react to prices and will endogenously operate his facility so as to maximize profit. Period. He will not have total flexibility to do so. He will be restricted by the size of his reservoir and the reserve requirements he places on his reservoir. He will be restricted by maximum rates of injection and maximum rates of withdrawal.

The recommended ArrowHead storage model has the advantage that:

- Gas supply and demand are balanced at every hub at every time point. For example, the monthly gas markets will equilibrate everywhere.

- Storage affects market price, and at the same time market price affects storage. The recommended model closes the loop, representing the fact that dispatch strategy depends on price and price depends on dispatch strategy. This is profound, and we are not aware that any other storage model embedded within an economic model has been able to accomplish it.

- It calculates inter-temporal opportunity cost of storing versus selling in every time point.

Some have wondered whether this is too cold-hearted and rational a model of the storage asset owner. Our long experience with this model suggests not; we have seen empirical evidence that owners of storage entitlements (meaning either storage field owners per se or third parties who have contracted for access to storage assets) do indeed maximize profit in exactly the fashion articulated in this section.

Certainly local distribution companies and public utility commissions (or their equivalents) around the world have traditionally viewed storage as insurance against high peak demands. Their reward and penalty structure forces them to behave this way. The recommended storage model represents this phenomenon very nicely. Using the reserve requirement, we represent the fact that there is a "nondiscretionary" portion of storage, a regulated portion of storage, that simulates the insurance aspect of storage that say a regulatory body in Europe would strive to enforce at time of peak. However, equally importantly, there is a residual piece, which is dispatched economically in pursuit of "buy low/sell high" profits. It is the combination of the two that impacts market prices and the value of the storage asset, and the recommended storage module represents that beautifully.

There are a large number of subtle details that must be present in the storage model EIA uses, and we have omitted those here. We have put forth the essence. We do not see how this type of model could be incorporated into a linear programming or complementarity formulation. They are going to omit endogenous representations of storage. The dimensionality alone of this model would dramatically exacerbate the dimensionality of those approaches.

12 ENDOGENOUS CAPACITY ADDITION

EIA absolutely must have a model that adds capacity endogenously and retires capacity endogenously. EIA has no realistic choice in the matter. The size and scope of the world hydrocarbon models EIA must have dictate that. And the requirement for endogenous capacity additions in your model occurs within every node in your model—transportation, refining (joint output), depletable resource, conversion, etc. It is simply not possible to externally estimate the capacity mix that will be in place 1, 5, 10, and 20 years into the future for many thousands of activities that comprise your world gas model or world oil model. Any such exogenous estimate is specious. It is so difficult for such a large scale problem that EIA has no choice but to use a model to do the heavy lifting on capacity additions.

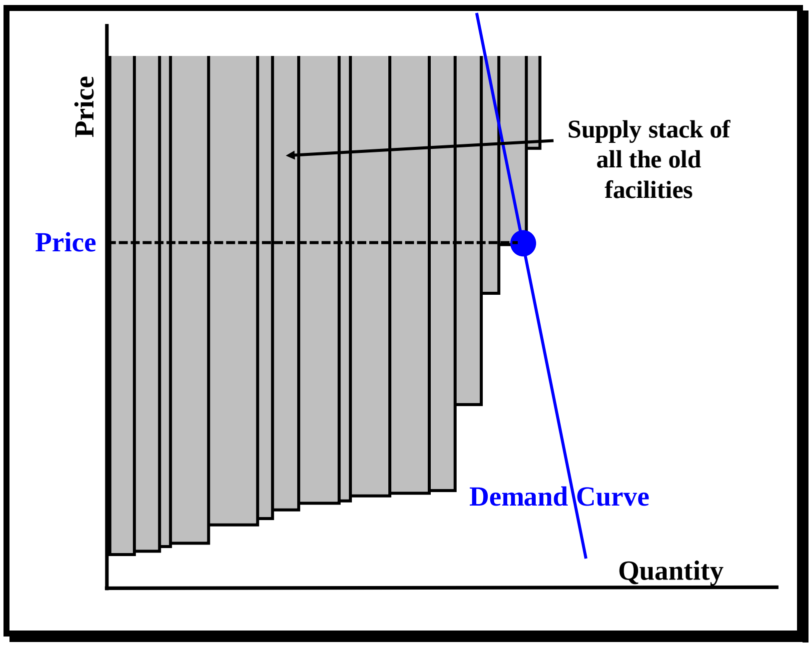
This section summarizes how endogenous capacity addition logic "thinks" and works. Our own capacity addition logic within our own ArrowHead system works this way (but with more detail and complexity). This section is not intended to be comprehensive on the subject but rather to give requisite insight. In particular, we show EIA the correct way to think about endogenous capacity additions to guide future development.

Consider a market at a point in time in the model, for example, German wholesale gas. The situation before any capacity addition appears as in Figure 67. There are a number of aspiring competing suppliers into that German wholesale market, each aspirant being represented by a "supply curve." The sum total of all supply curves collectively comprises a "supply stack," a merit order in power plant parlance, into the German market. The height of each of the supply curves represents the variable cost of the aspiring supplier in the market. The width of each of the supply curves represents the capacity in place for that particular aspiring supply into that market. The notion of supply curves coming into Germany is scarcely a foreign concept; we have emphasized that notion throughout this paper.

The situation in Figure 67 pertains in every market everywhere in the world, not only German wholesale gas. Every market sees a collection of existing supply sources of different sizes and of different costs aspiring to sell gas into that market. The capital cost and fixed cost of all such capacity is "sunk" once the capacity is in place. In addition, and it is crucially important to note, there is a demand curve in that market as well (depicted by the blue line). The demand curve crosses through the existing supply stack at a point depicted by the blue dot in Figure 67. In our example, that crossing point represents the price in the German market in terms of its vertical position and the quantity that flows into the German market in terms of its horizontal position. This blue dot is the situation that would occur in the German market if there were no new capacity added, i.e., if all plants in the German market were exogenously given, old plants. This is the type of model EIA has always had as we understand it, a model that requires capacity to be exogenously assumed and given, not endogenously determined as we are about to describe.

We now ask the question: What would happen if one added very small amount of capacity from the left side of Figure 67? (The capacity would almost always enter the supply stack at the left because the variable cost of new capacity and the thermal efficiency of new capacity would be lower than most old vintages and therefore its variable cost would be lower. Whether or not the entry is precisely from the left is irrelevant to our argument, but it is convenient to draw the picture if one assumes entry from the left. (Our algorithm does not assume that.) Figure 68 shows a small amount of capacity (the horizontal width of the green rectangle at the left) added from the left. The horizontal width of the green rectangle at the left is the amount of capacity added. The margin capture of such capacity addition is equal to the margin (the vertical height of that green rectangle at the left) times the capacity added (the horizontal height of that green rectangle at the left). This notion becomes quite convenient. As the producer adds a small amount of capacity to the left of the supply stack, the margin capture by that small amount of new capacity is equal to the horizontal width of the green rectangle (capacity) times the vertical width of the green rectangle (margin) in Figure 68. Very importantly, as that small amount of capacity is added at the left, the entire supply stack of old, existing plants is displaced and forced to the right.

Figure 67: Supply Stack of All the Existing Plants



If the capacity is small in magnitude as Figure 68, the supply stack is not displaced very far to the right, the market clearing price is not decreased very much, and the quantity is not increased very much. However, as is clear in the curve, the margin capture of this small amount of capacity that is added is positive—not zero and not negative. The green area is positive for the small magnitude of capacity added. Margin capture is margin times capacity, and margin times capacity is positive in this case. (We will discuss the capital cost of adding this small amount capacity below to complete the picture, for now, we focus on margin capture). The point to be made is that positive margin capture would occur if the producer added capacity in this market at the left of the supply stack.

Now suppose that the producer decided to add even more capacity at the left of the supply stack as shown in Figure 69. The supply stack comprised of all the old, previously existing plants is displaced further to the right by the new plants added from the left. The amount added is the horizontal width of the green rectangle. But something has started to happen. As the supply stack is displaced to the right, the demand curve is cutting through the displaced supply stack at a lower price. The price in Germany has begun to erode by the entry of new capacity into the market. The market is more strongly supplied relative to the situation before any capacity is added, and that additional supply starts to erode the price. The market clearing price drops from its old level of the blue dot to the new, lower level of the green dot because of the increased supply in Germany resulting from the new capacity.

Figure 68: A Small Amount of New Capacity Is Added to the Supply Stack of All the Existing Plants

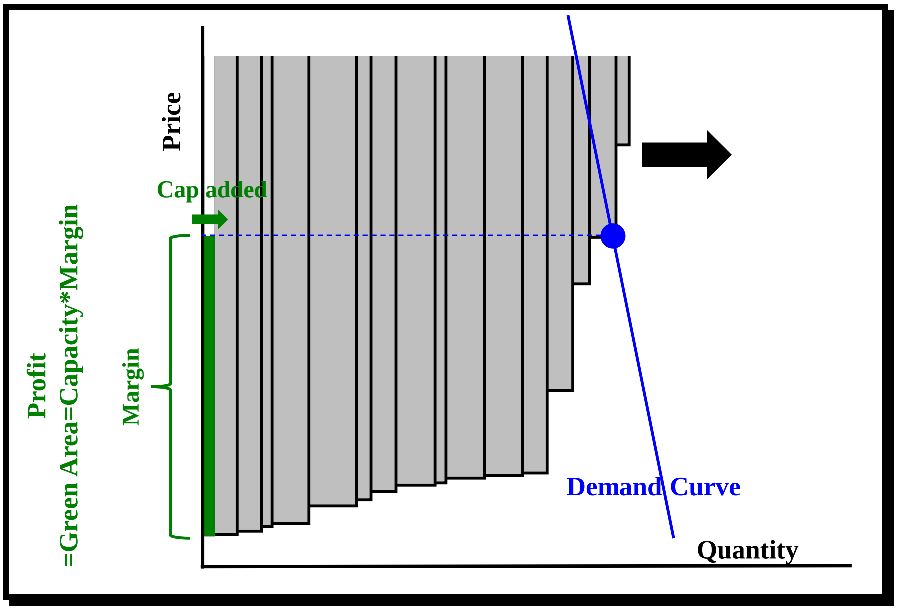
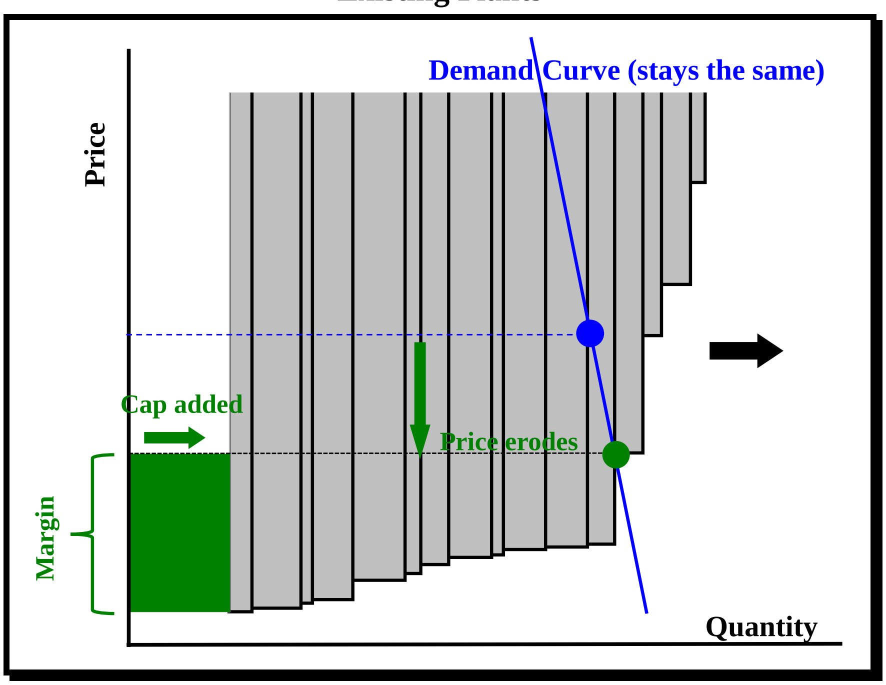


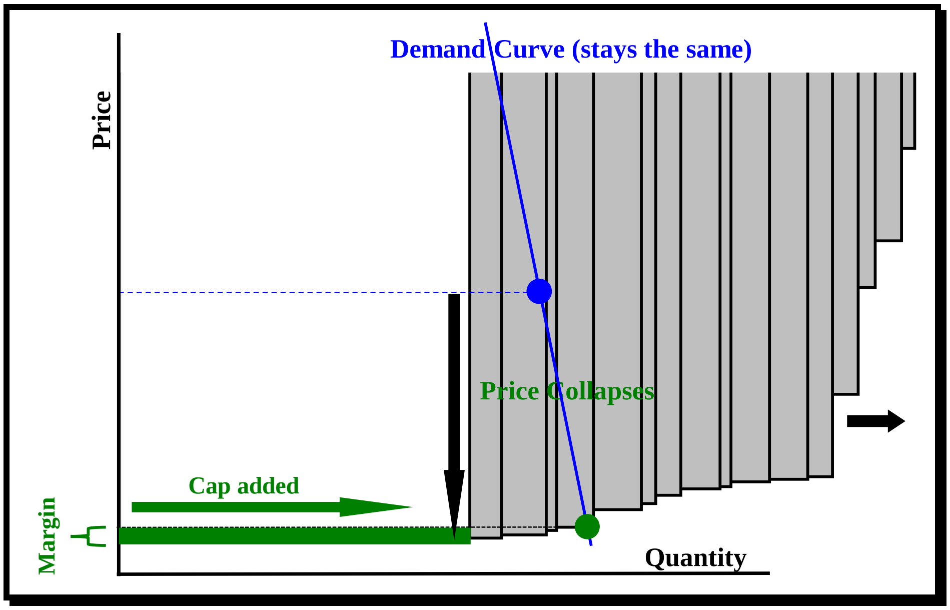
Figure 69: A Larger Amount of New Capacity Is Added to the Supply Stack of All the Existing Plants



If we calculate the margin capture that results from this larger volume of capacity added in the German market, it is the horizontal width of the green rectangle (the capacity added) times the vertical height of the green rectangle (the margin), i.e., the area of the green rectangle in Figure 69. The margin is positive, but it has begun to "scrunch" down in height because the market price in Germany is beginning to erode as the new capacity enters. This is the absolutely crucial dimension of capacity addition. The producer MUST take account of the change in market price that the entering new capacity induces. The margin capture in Figure 69 (the area of the green rectangle in Figure 69) is higher than the margin capture in Figure 68 (the area of the green rectangle in Figure 68). The market is able to absorb the higher capacity addition in Figure 69 than in Figure 68, providing a higher margin when capacity addition is higher.

If we add a still higher, and in fact very large magnitude of capacity as in Figure 70, the situation changes markedly. The supply stack comprised of all the old plants moves even further to the right, but the demand curve, of course, does not shift. Notice that the old blue price falls, very markedly, all the way to the badly depressed green price, i.e., all the way to the variable operating cost of one of the best supply options within the German market. The margin capture remains the massive amount of capacity added (the horizontal width of the green rectangle) times the margin (which has been depressed almost all the way to zero because the market is so highly overbuilt. A large magnitude of capacity addition times a zero margin would be a zero margin capture. The margin capture has gone to zero in this case of massive overbuilding. The German market could not absorb this magnitude of new capacity without crashing the price all the way down to variable cost.

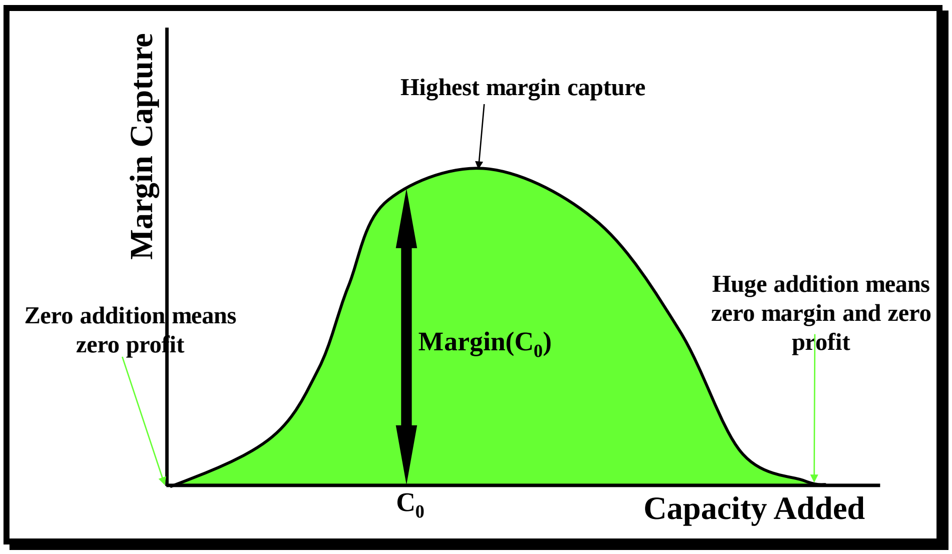
Figure 70: A Very Large Amount of New Capacity Is Added to the Supply Stack of All the Existing Plants



What does Figure 70 mean? It means if that if the producer adds a large enough magnitude of capacity, the margin capture goes to zero because the producer has completely overbuilt the market and killed the price in that market, driving it all the way down to variable cost.

What do Figure 68, Figure 69, and Figure 70 imply collectively? They imply that that the margin capture from capacity addition looks like a "Laffer Curve." The margin capture is zero if zero capacity is added (capacity times margin is zero because capacity is zero). The margin capture is zero if a huge amount of capacity is added (capacity times margin is zero because margin is driven to zero). Figure 71 illustrates. The notion of a Laffer curve is evident in the diagram—margin capture starts at zero, rises to a maximum point, and then falls back to zero. There is going to be a point at which margin capture is maximum, taking full account of the depression in market price that new capacity entry induces. Ignoring, as we have thus far, the issue of capital cost, what level of capacity would make the producer the most money in terms of margin capture? To ask the question is to answer it—the highest point!

Figure 71: "Laffer Curve" for Profitability from Addition of Capacity



Now we simply superimpose the notion of the capital cost of entry. The capital cost of entry is linear in capacity. Each unit of capacity costs the same per unit of capacity. The total capital cost curve is, therefore, just an upward tilting line as shown in Figure 72.

Once we have the curves in Figure 72, we can subtract the blue capital cost curve from the green margin capture curve to calculate the true profit curve from entry. Notice there are domains in Figure 72 where the capital cost curve (blue) is below the margin capture curve (green). Those represent capacity addition levels that are not profitable, and the producer would not be expected to implement them. There are also domains where the capital cost curve is below the margin capture curve. Those are places when capacity addition levels are profitable. Subtraction of the capital cost curve in Figure 72 from the margin capture curve in Figure 72 yields the true Laffer curve for the true, total profitability of capacity addition. That curve appears in Figure 73. The curve in Figure 73 is the curve EIA must use to predict addition of capacity. The model adds capacity that yields the point of maximum profit.

In a nutshell, the logic in Figure 73 is how EIA must represent endogenous capacity addition. The ArrowHead capacity addition algorithm does this.

Figure 72: “Laffer Curve” for Margin Capture from Addition of Capacity

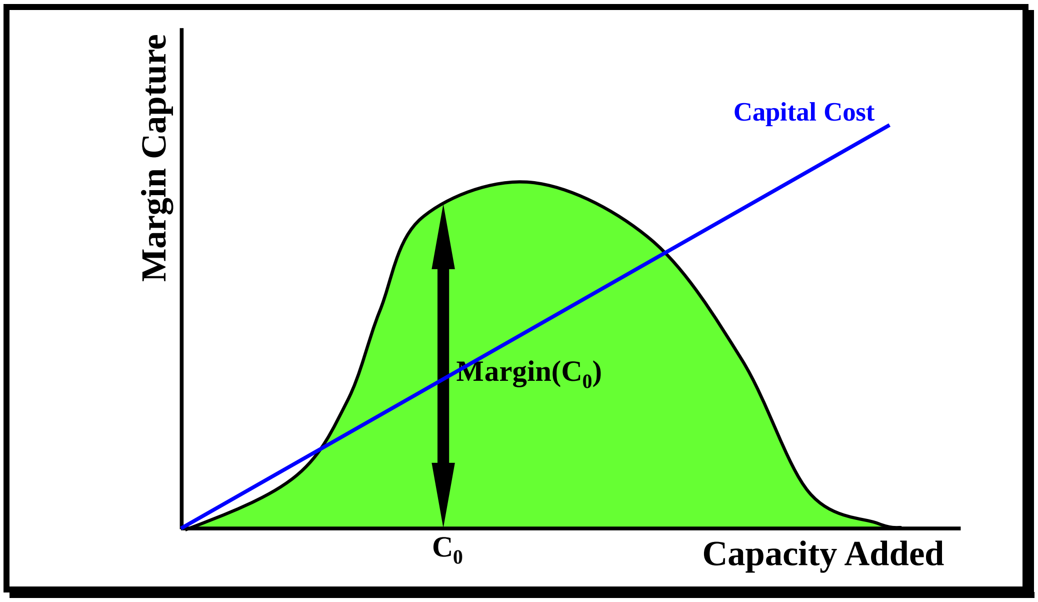
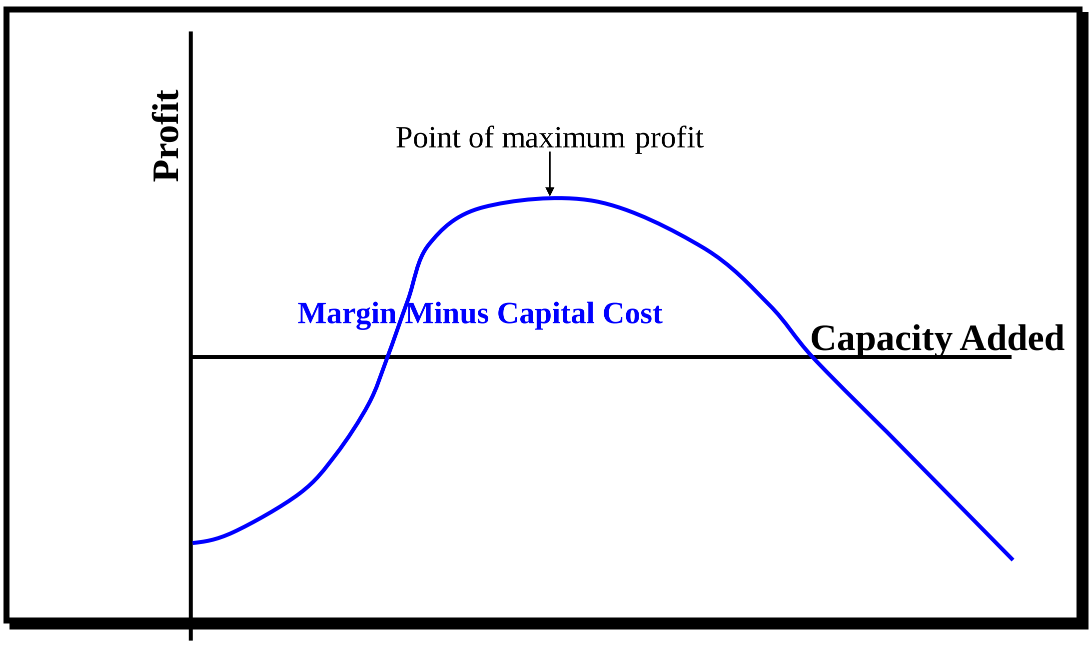


Figure 73: “Laffer Curve” for Profitability from Addition of Capacity



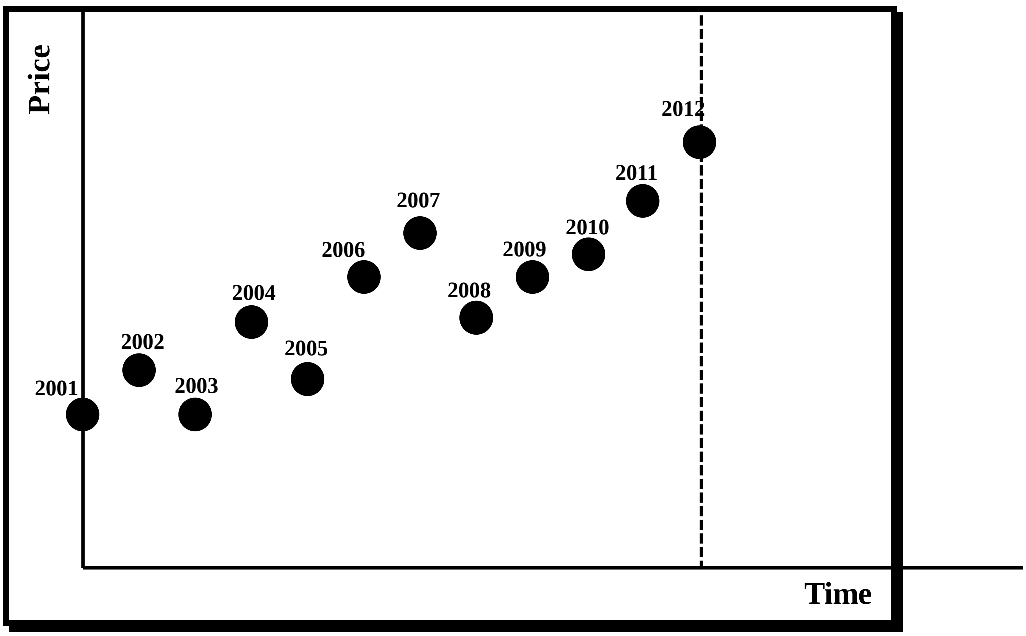
13 MATCHING MODEL RESULTS TO HISTORY

Matching model results to history has been one of the most popular suggestions by model dilettantes. One can do it, and we have done it, with our own models. However, history matching or backcasting adds precious little value to or confidence in models or their predictability. We make it a point to charge clients if they want to do it; if they ascribe any value to it, they pay for it. This section puts forth some important issues related to model backcasting.

13.1 The Problem with Historical Validation Is Incomplete Data

Following is the stereotypical approach to historical validation. Someone will provide a time history of spot prices at Henry Hub. The plot will look something like that in Figure 74. There will be a plot of observed historical prices at various points in time (assumed to be annual in this example).[38][39]

Figure 74: Historical Prices



It is asserted that this ought to be enough. You ought to be able to start your model in the year 2001 in Figure 74, and you ought to be able to see your model "walk right through the forward points" thereafter. Start your model in 2001 with what you think are 2001 initial conditions, and run it forward in time from then. If your model is any darn good, it is argued, your model will run right through the points 2002, 2003, 2004, …, 2012.

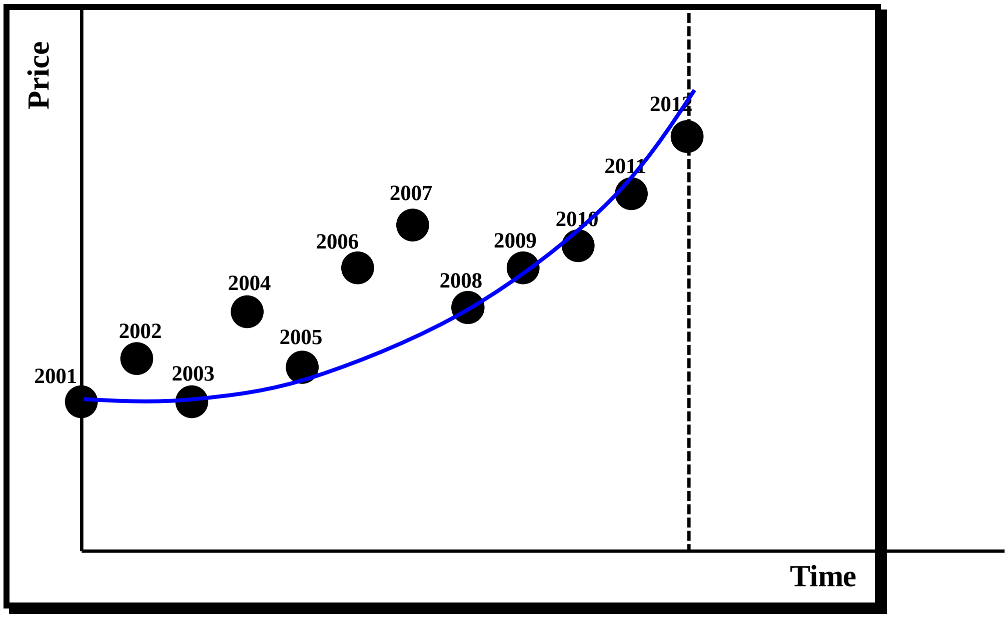
[38] We are going to omit the economic notion of "identification" here, which is an extremely important yet subtle issue. "Identification" asks the question what one is actually observing in these historical time series. The "identification" discussion is for another day.

[39] We will not but not pursue the notion that historical price observations themselves can be highly inaccurate. They might not have been at all observable. They might have been guessed by an analyst. For the moment, we will assume that they are accurate and correct, but that is a big stretch. Historical data is replete with inaccuary in practice.

Is such an assertion really realistic? Of course not. To see why, let us display a couple of diagrams and then judge whether one is better than the other.

Would a model run relative to history such as that in Figure 75 be a convincing historical validation? It certainly seems to consider all the lower price points. It simultaneously seems to ignore the higher price points. Intuitively, it does not seem to be weighting all the historical points equally. Interestingly, if by historical happenstance the data set had serendipitously omitted some or all of the higher price points, the blue curve would look fantastic to many people.

Figure 75: Case 1—Through the Lower Points



How about the fit shown in Figure 76. This is an absolutely "perfect" fit. The model projection (the blue line) goes right through every historical point. Zero error. Zero deviation. Perfectimundo! What could be more convincing than this? The model replicates history exactly. When people see something like this, the model and all the modelers actually become **<u>less</u>** credible, not more credible. "If it looks too good to be true, it probably is." Dr. William Hogan was once reputed to have quipped: "If you insist, I can fit an electrocardiogram with my model. I know where all the levers are." The gist of the comment was prescient. What earthly good would that type of fitting do for anyone? Would people bow down to a model that was this uncanny at predicting the past? Wow! Perfect predictor of history! Never missed a historical point. We have witnessed a lot of historical validations, particularly with linear programming formulations, that attempt this type of validation. None of them has worked, most being downright dishonest.

There are a couple more prescient points. Suppose we said "We would like the forecast from the historical validation to behave like a regression model. We would like the forecast to be a prediction such as statistical/econometric people are wont to do. Its root-mean-square error of predicted versus actual summed across the years would be "minimum." This is actually preposterous on its face. What does the classical regression model assume? It assumes that the historical data points were generated from a univariate normal error function of predicted minus actual. It posits a functional form and fits in such a fashion that the root mean squared error is minimum, as illustrated in Figure 77.

Figure 76: Case 2—The "Perfect" Fit

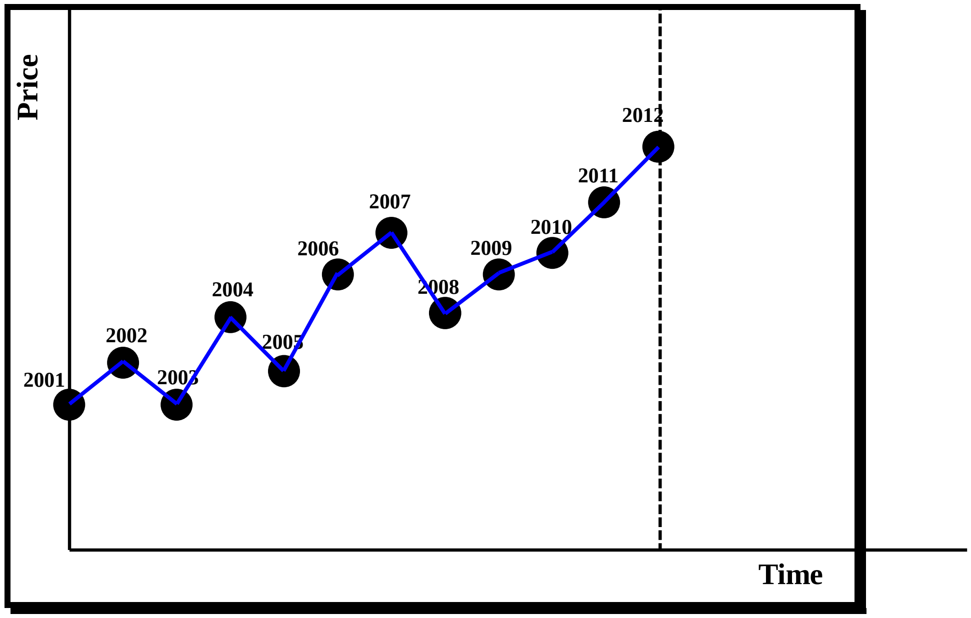
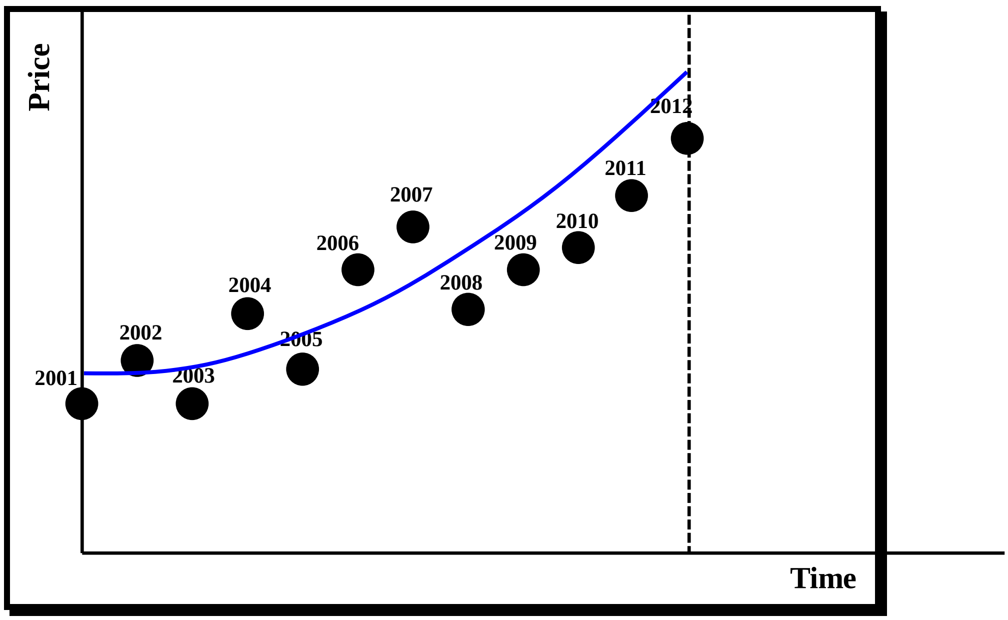


Figure 77: Case 3—Minimum RMS Fit

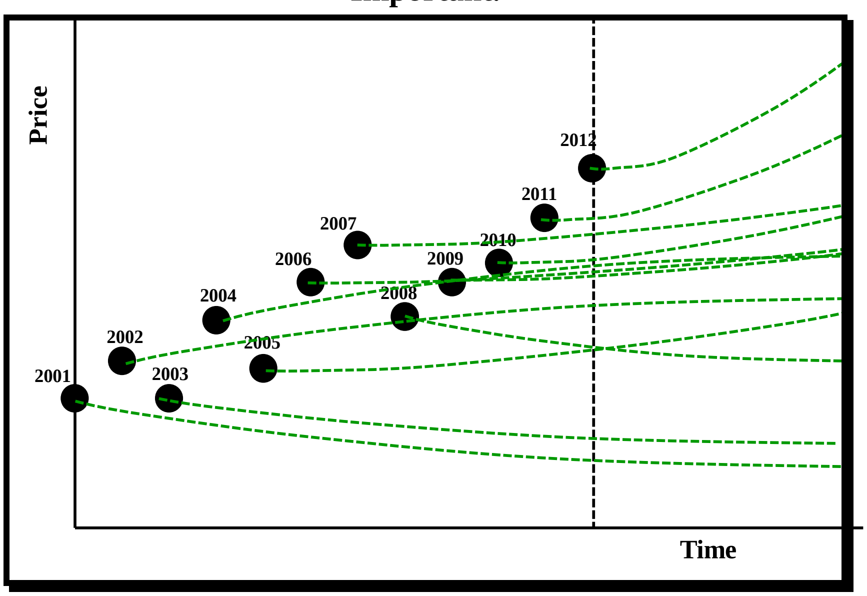


There is another very serious problem with statistical "fit" sorts of models with respect to historical data. For statistical methods to work, we are fully aware that the data points must have been generated by a (homoscedastic) random error function. We know that is and can **never** be true with an economic model. An economic model is a **game theory problem**! There is absolutely nothing random about it. The reality that the historical price-quantity observations were generated by a multiperson game theory problem and not a random process. There simply is no basis whatsoever to use a statistical process on historical data. It was not generated using a random process. It, therefore, cannot possibly satisfy the sufficiency conditions for statistics.

The final challenge, and really the *coup de grace* for historical validation, is that historical data is and never can be complete. We know that economic models have price expectations within them, just as the real world has price expectations within it. Respectable economic models have all their agents engaged in price expectation. In a historical data set, there is no way to observe long since gone expectations that pertained during the historical years, and yet such expectations had a huge impact on what happened in 2001, 2002, 2003, …, 2012 in terms of installed capacity, price, quantity, etc. There is an insurmountable absence of requisite historical data, namely expectations of the future that existed back then. It can never be gathered.

Why is this important? No real world resource producer and no model drills into the “spot” market. Models and markets drill into anticipated forward prices, and anticipations are somewhat accurate. No model does or can weld pipe into a “spot” market. Models and markets build pipe into anticipated forward origin and destination prices. No model does or can build refineries or power plants into “spot” markets. Refineries and power plants install capacity into forward crude, product, power, and fuel prices. Forward expectations are ubiquitous in markets. As Figure 78 emphasizes, every single, solitary historical year had, at the time when producers were living in that year, a set of price expectations that were strongly driving producer and consumer decisions. And those price expectations were both crucially important and fleetingly ephemeral; they flitted away when the clock changed to the next year. There is no way to resurrect them now (whether or not there was a way to observe them even back then). The green dotted curves emanating forward from each historical point, which represent expectations forward in time from that historical point, emphasize that. Every historical time point had not only the price at that time point but in fact a full forward time vector of price expectations. No one today can see past price expectations.

Figure 78: Expectations from Each Period in the Past Unobserved, Unobservable, and Important.



Alas, there is absolutely no way in historical data to know or even begin to conjecture what forward expectations were at each point during the historical period. This is a huge, glaring, mortal omission that cannot be overcome in historical validation and backcasting. What in the world were those green dotted curves in Figure 78? No one knows. No one will ever know. And yet they were MAJOR DRIVERS of what happened during those years and the subsequent years that ensued. The inability to gather data on historical expectations all but dooms historical validation. There is a classical, and irreparable, missing data problem with historical backcasting.

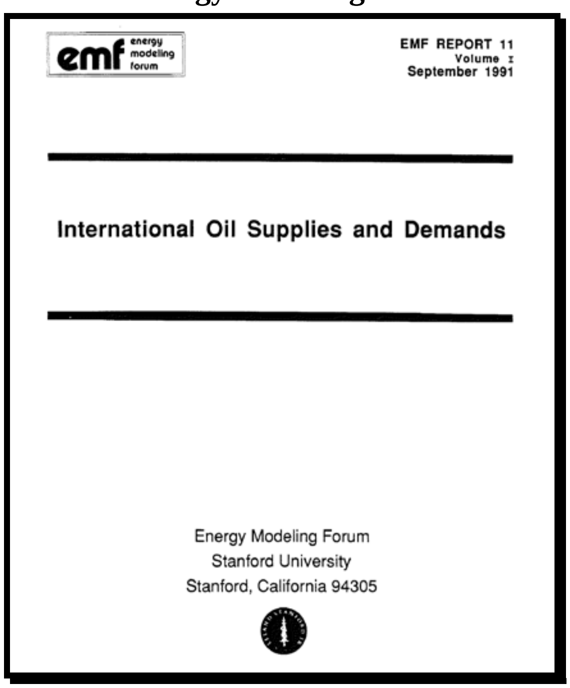
In summary, there is no "right" answer, and there is no convincing result. History matching does absolutely nothing to underscore future accuracy. How many times have we seen a model match history fairly nicely but pratfall badly in almost any prediction it ever subsequently makes? History matching proves nothing.

13.2 A Historical Validation Example Using Our Own Model

We will show a quick analysis from our own history. This is, in our opinion, about the only backcasting that might be relevant or useful. The beauty of it is that the model truly started some years ago in a distant historical year, and there is no question about prices or expectations during the intervening years. There is no "cheating." Those expectations were in the model. The model was actually a pure forward projection when it was done.

The year was 1990. It was Stanford University Energy Modeling Forum Number 11: World Oil. See Figure 79. We were just coming out of the very low price era of the 1980s, and oil price was on the rise. We and all the other participants were looking at models that were built in the 1980s using our technology and run in approximately 1990.

Figure 79: Energy Modeling Forum Number 11



emf energy modeling forum

EMF REPORT 11
Volume I
September 1991

International Oil Supplies and Demands

Energy Modeling Forum
Stanford University
Stanford, California 94305

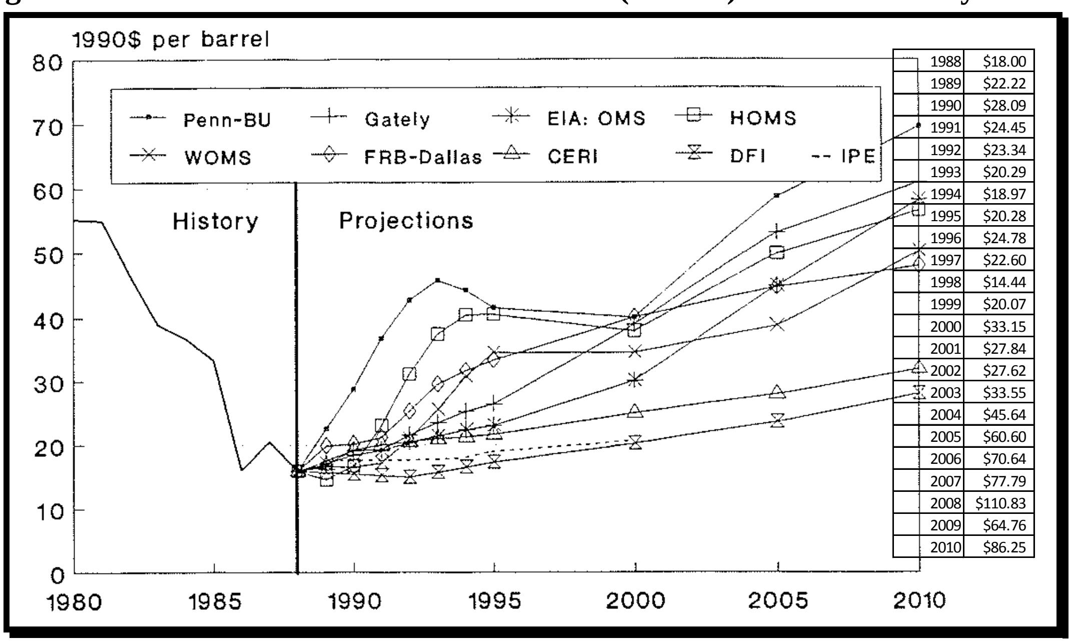
There were twelve (12) energy models participating in the study of world oil were those, as published in the foregoing report, listed in Figure 80.

Figure 80: Twelve Participating Models

Model	Working Group Contact*
EIA:OMS	Mark Rodekohr, Energy Information Administration
IPE	Nazli Choucri, Massachusetts Institute of Technology
ETA-MACRO	Alan Manne, Stanford
WOMS	Nicholas Baldwin, PowerGen, U.K.
CERI	Anthony Reinsch, Canadian Energy Research Institute
HOMS	William Hogan, Harvard, and Paul Leiby, Oak Ridge National Laboratory
FRB-Dallas	Stephen P.A. Brown, Federal Reserve Bank of Dallas
DFI-CEC	Dale Nesbitt, Decision Focus, Inc.
BP America	E. Lakis Vouyoukas, British Petroleum
Gately	Dermot Gately, New York University
Penn-BU	Peter Pauly, University of Pennsylvania and University of Toronto, and Robert Kaufmann, Boston University

The twelve modeling groups had been running scenarios and conducting analysis for some two years prior to the final meeting. The final results from the twelve models were those in Figure 81, taken directly from the referenced 1990 report. Notice that there were two groups of results. There was a "low" group of results consisting of DFI (Dale Nesbitt) and the Canadian Energy Research Institute (Tony Reinsch). The Nesbitt result was the lowest of all, and by a good margin. It was the lowest line in Figure 81.

Figure 81: There Were Two Clusters—DFI (Nesbitt)/CERI and Everyone Else

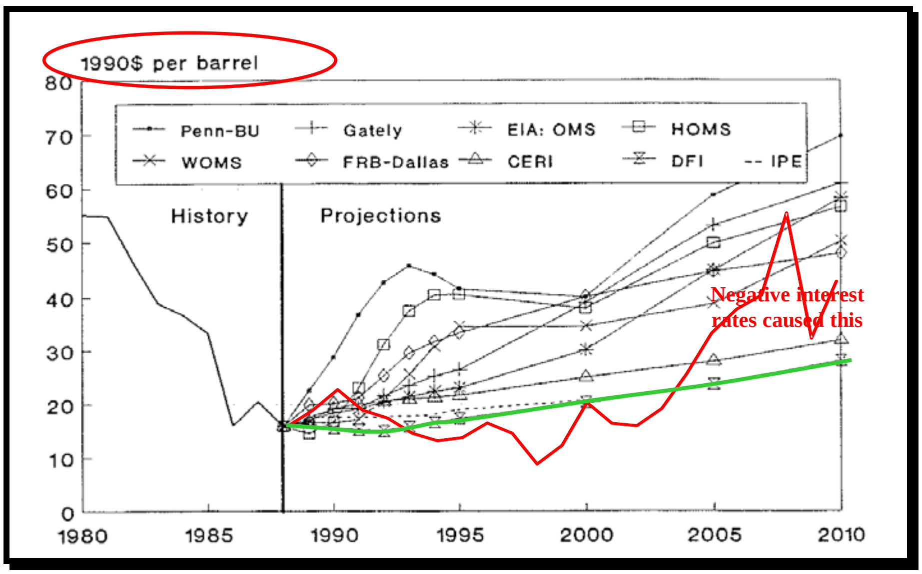


We recall (and Hill Huntington of Stanford will vouch) the final meeting in Banff, Alberta in September of 1990. The late Dr. Campbell Watkins said: "There are two clusters of results. 10 are really, really high. Two are substantially lower than everyone else. We have pilloried Nesbitt and

Reinsch for 18 months now. Let's have a vote, a plebiscite, of the members which of the two clusters we think is right, sort of a voter's consensus." Amazingly, the vote was something like 25-1 in favor of the low forecasts, including all the other modelers except for Dr. Dermot Gately. (Dr. Gately said he just couldn't vote against his own model, which was among the highest.) It was amazing; the modelers mostly voted against their own models, including EIA. I will never forget that vote (nor the Canadian Rockies).

And what actually occurred? It turned out that DFI (Dale Nesbitt) was spectacular relative to actual. The red curve in Figure 82 is the actual, true oil price that actually occurred according to EIA. You will notice that DFI (Dale Nesbitt) was far and away the best for almost twenty years forward. The final five years were far into the murky future in 1990 when these forecasts were prepared. (Needleless to say, no one in 1990 forecast the financial crisis of 2008 and its negative interest rates.) DFI was actually quite uncannily accurate. (Ponder the temporal scale of this diagram; it is rather amazing.)

Figure 82: Actual Oil Price versus All the Forecasts



This is not unlike the EIA retrospective that was published by EIA a few years ago. That type of thing can be somewhat useful although never definitive. That is the type of thing we recommend doing—information history matching by models of the historical day. Our models have fared quite well under informal historical validation.

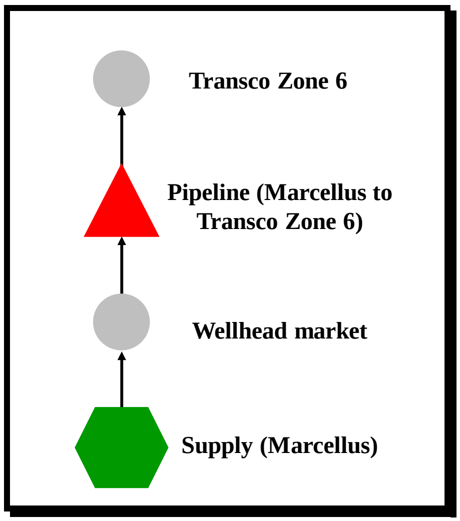
Returning to Figure 82, whose forecasts are better? The answer was pretty clear. Whose forecasts would have supported correct decision making better? The answer is pretty clear. Whose forecasts would have caused their clients to lose money? The answer to that is pretty clear too. Why did we do so well in retrospect? It was because my company at the time DFI had an agent-based, network-oriented, structural model of world oil supply, transportation, and demand. Others had linear programming, nonlinear programming, or "reaction function" models, which simply do not work well. DFI didn't have a bunch of "reaction functions" and other structural "tumors" in our model that biased the results high. It seems there is no substitute for supply, transportation, refining, and demand and integrating them all together. Michael Lynch's work has shed light on biases in models owing to model structure (e.g., "hockey stick" functional forms). We are not going to reference or analyze that work here, but we recall some very nice analysis by him identifying bias in models owing to their structural forms, which were not network equilibrium models.

Now that we have put this forward, does this historical validation hang the stars in the sky? Convince everyone they should avoid all these other model and go with DFI's because it was so prescient? We believe so, but many others immediately begin to raise "Yeah but's." Yeah but… Yeah but… Yeah but… Yeah but… The chorus of "Yeah but's" would be deafening. People may clamor for or insist on this type of backcasting, but in the final analysis people are not convinced by historical validation. They have "Yeah but's" coming out their ears. That is reality. Talk is cheap. That is precisely why we charge people to do it. If it is important to them, they will pay for it. If not, they will move on and make better decisions. We recommend that EIA de-emphasize it if you can.

14 VERTICAL INTEGRATION

The question has arisen: Does EIA need a vertically integrated model that coordinates the decisions of agents, i.e., that forces simultaneous decisions by two or more nodal agents? We address the question in the context of Figure 83. The figure postulates the supply of gas from Marcellus shale and the subsequent construction of a pipeline from the Marcellus to Transco Zone 6 in order to access New Jersey and New York markets. Do we need to "force" a resource development decision in the green supply node to be timed and coordinated with a pipeline construction decision in the red pipeline node? After all, no one would build a pipeline before the gas is proved and ready to go, and no one would build a pipeline before the gas reserves were proved and ready to carry. Can you leave it to the so-called serendipity of a decentralized model in which the resource producer develops the resource whenever and however he wants, and the pipeline constructs the pipe whenever and however he wants?

Figure 83: Coordinated Development of Marcellus Supply with Outbound Pipe



In a modeling sense, the question is can we create a "meta-node" that encompasses a vertically integrated resource and pipeline decision as presaged in Figure 84? The answer is, yes you can. The answer is also that you do not need to.

That vertically integrated node would be a meta node as indicated in Figure 85. Sure, you can combine the capital and operating cost serially of the resource and pipeline, and you can let the model decide the prospective entry of the vertically integrated whole. It is very easy to combine the resource and transmission node into a single vertically integrated, single decision project. Just combine the capital cost, the operating cost, the decline curve, and the various parameters of the supply node and the pipeline node and attach the vertically integrated node into the target market you want to consider and Voila! you have it.

Figure 84: Can We Create and Do We Need a Vertically Integrated Model of Supply and Pipe?

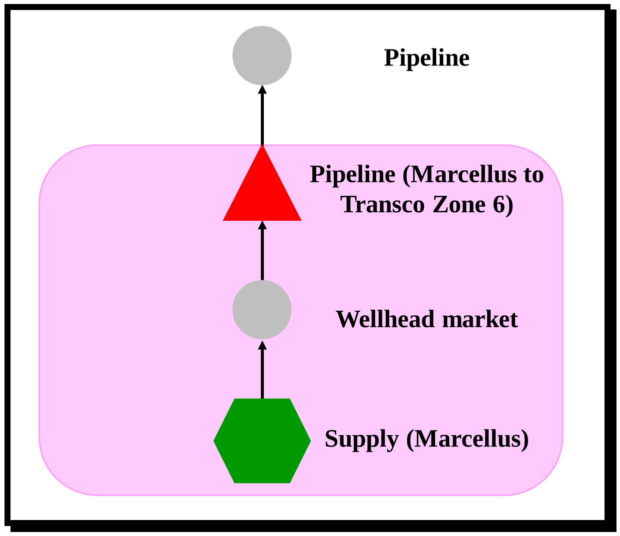
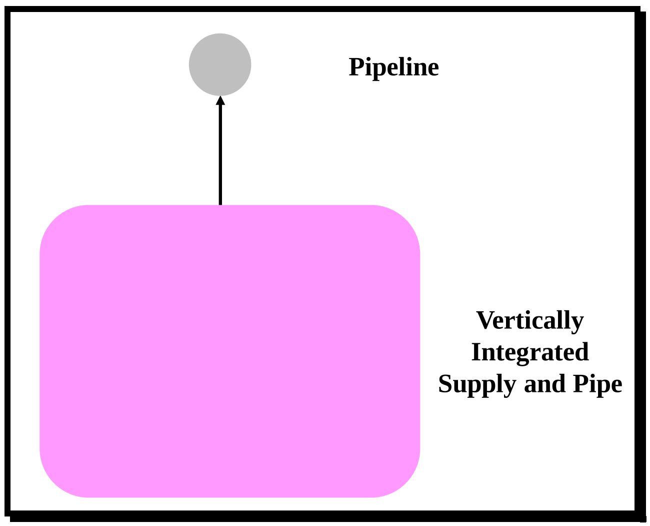


Figure 85: A Single Vertically Integrated Production and Pipeline Model



The reason EIA does not want to do this is twofold. Using the decentralized configuration in Figure 83, the model will determine what the profitable and competitive activities are. One would expect things predicted in a decentralized fashion to actually come to pass in a market. Secondly, in the context of Figure 83, there are invariably going to be other competitive uses for Marcellus wellhead gas, not necessarily any particular vertically integrated project one wishes to posit. It is never a good idea to "assume" that a vertically integrated supply and pipeline meta-project will stay vertically integrated. One person's vertically integrated project is another person's negotiating opportunity. New supply volume can enter, and new pipe and destinations can enter. You want your model to consider those options, and the way to do so is to make sure it is not vertically integrated.

15 TEMPORALITY AND TIME PERIOD CONVENTIONS

The issue of time points and time horizons is an important and time-consuming issue. This section summarizes the timing and time point conventions that the models of transportation, logistics, and in fact all economic activities should follow.

We recommend that the forward time grid in EIA's modeling system be completely and totally variable and arbitrary. **The time grid has to be an input to the model**. That means that the model has to accept any time grid simply by typing in new time points. The time grid has to be a piece of data to the model, just like all numerical data. In particular, we recommend that the forward time grid have the structure in Figure 86. Notice that there are time points, prices, flowing quantities, and inter-period discount factors as follows:

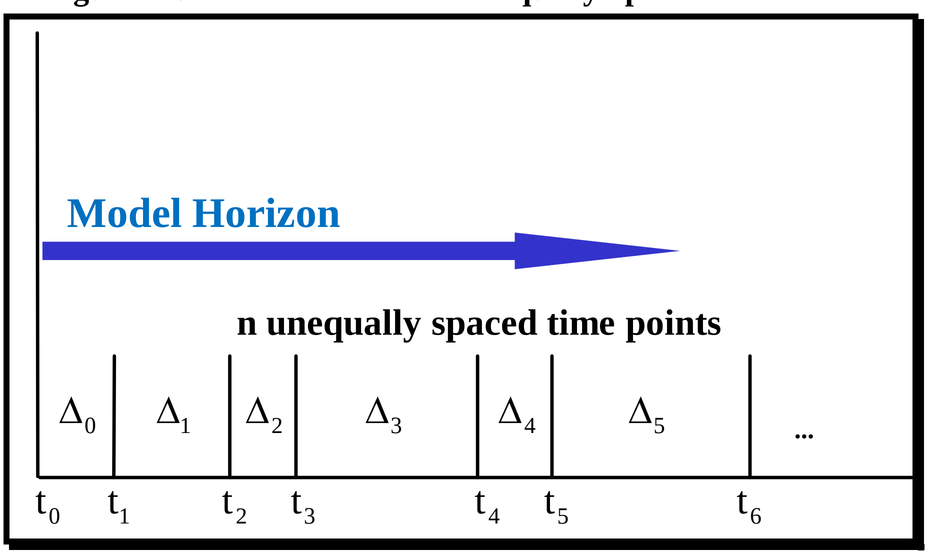
$$\text{Time Points: } t_0, t_1, t_2, \ldots, t_N$$

$$\text{Prices: } p_{t_0}, p_{t_1}, p_{t_2}, \ldots, p_{t_N}$$

$$\text{Quantities: } q_{t_0}, q_{t_1}, q_{t_2}, \ldots, q_{t_N}$$

$$\text{Interperiod Discount Factors: } D_{t_0}, D_{t_1}, D_{t_2}, \ldots, D_{t_N}$$

Figure 86: Time Grid with Unequally Spaced Time Points



This forward time point structure should be embedded and institutionalized in every node in the model (in the case of nodal microeconomic modeling) and every equation in the model.

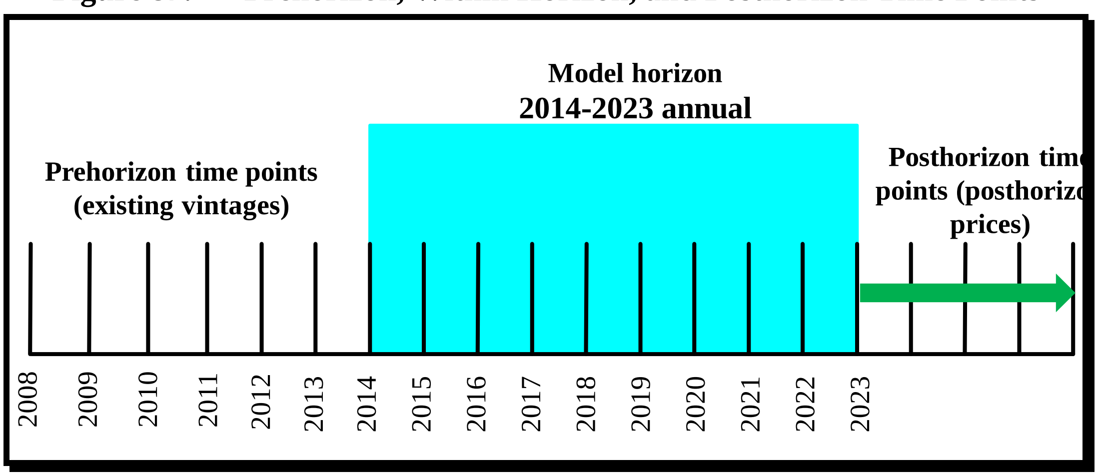
As EIA considers the dynamic environment of its model, consider that the inter-temporal extension of Samuelsonian consumers' surplus does not generalize particularly easily to the dynamic case. This would have to be done with the most extreme care; otherwise, your global welfare maximization formulation will not be isomorphic to a dynamic equilibrium, and your solution will be utterly arbitrary, meaningless, and in fact dishonest. It will assume that all your agents march in simultaneous, collusive lockstep to whatever measure of dynamic social welfare you choose either advertently or inadvertently. That has been a difficult hill to climb for monolithic global welfare maximization. Dynamic consumers' surplus requires a dynamically integrable demand function.

You not only need the foregoing "within model horizon" time point structure in Figure 86, you also need a premodel time point structure to account for all the pre-existing vintages of plants and wells that are in place at the beginning of the model horizon and have the potential to produce during the model horizon. You also need to make assumptions about what happens after the model horizon because you do not want the agents in your model to "game" the final time point in the model solution. You do not want them anticipating, quite incorrectly, that the world as they know it ends at the end of the model horizon and there are no prices or economic values for anything after the end of the model time horizon.

For shorter time horizon models, you will not allow endogenous capacity addition of transportation, storage, consumption, resource, or economic production. You will exogenously input the vintages of historical capacity and any assumed capacity entries during the model horizon. For longer time horizon models, you will allow endogenous capacity addition at any and every model time horizon point.

There is one more very important consideration for the timeline. Figure 87 illustrates that there are actually three timelines that EIA must consider in its model horizon (and they all must be definable and enterable into a common modeling software structure). As indicated in the figure, EIA must consider prehorizon time points. There is an age distribution of plants in place, built before the model horizon, and many of those plants are destined to terminate their operating lives during the model horizon. EIA needs to have a fairly lengthy prehorizon set of time points for each process to characterize the temporality and capacity of previously existing capacity already in place. (ArrowHead has a general prehorizon set of time points in its nodal models.) Capacities in place should be associated with premodel time points when they were built so that they will be retired at appropriate time points within the model horizon commensurate with their operating lives.

Figure 87: Prehorizon, Within Horizon, and Posthorizon Time Points



We see in Figure 87 the model horizon, designed in this illustration 2014 through 2013 in annual increments. Such a model would be a ten-year horizon with accounting for capacity that was in place before the model horizon and the time at which it was built and the time it might be retired. We also see in the model a series of posthorizon time points. These time points are used to input what are termed "transversality conditions" in optimal control problems (i.e., agent profit maximization problems). These transversality conditions posit prices, operating costs, and profit margins for plants that will be built during the model horizon but will operate and generate profits (and therefore generate present value of profits) past the end of the model horizon. EIA's models must take explicit account of the three time horizons:

1. Prehorizon time points and prehorizon vintages of capacity
2. Within horizon time points, including operating and capacity addition
3. Post horizon time points, which help to determine capacity addition for plants built within the model horizon but operating after the time horizon

The time horizon must be completely and totally arbitrary. EIA must be able to specify:

- the number of forward time points that comprise the model horizon.
- the intervals between those within horizon time points.
- the number of prehorizon time points.
- the intervals between those prehorizon time points.
- the number of posthorizon time points.
- the intervals between those posthorizon time points.

The time intervals must be completely arbitrary and completely user specified—hours, days, weeks, months, years, multi-year time intervals, etc. The time grid should be built intrinsically into the node logic. (We have built an hourly model of electricity storage (batteries and capacitors), and we have built a multiyear model of long-term natural gas, all with identical node and solution software. That is the ideal.)

16 DATA MANAGEMENT

Assuming EIA secures a network microeconomic equilibrium model, the data management issue becomes one of storing and delivering the coefficients of the production functions, variable costs, fixed costs, capital costs, installed capacities, and book and tax regimes and that pertain to every node in the model. Assuming the implementation is regionally hierarchical, that means your data management system is in the form of a standard hierarchical database. That is eminently convenient. It says you can store your data in a standard database product and deliver it to your model.

The data creation and editing functions, all of which must be executable by Excel, are fairly simple:

- Single node comprehensive input. You must be able to retrieve an Excel spreadsheet that is automatically populated with all the data (and that means 100 percent of the data) for a selected node. You must be able to modify that spreadsheet, and you must be able to have that spreadsheet reinserted into the model automatically.

- Multi-node single variable input. You must be able to retrieve an Excel spreadsheet that contains a given variable (e.g., capital cost) for every node in the model. You must be able to modify that spreadsheet, and you must be able to have that spreadsheet reinserted into the model automatically.

- Pivot table input for multiple nodes and variables. You must be able to retrieve an Excel spreadsheet that contains two or more selected variables (e.g., capital cost, variable cost) for every node in the model. You must be able to modify that spreadsheet, and you must be able to have that spreadsheet reinserted into the model automatically. This is an extension of the single variable case with logic that puts two or more variables onto the same page.

Your EIA hierarchical data management system can be implemented in a database product without any difficulty at all. Your EIA database can be a series of spreadsheets as well. Both are interfaceable with the network structure.

17 RECOMMENDATIONS FOR IMPLEMENTATION

As EIA implements a world gas or world oil model complete with transportation and storage logistics (or an electricity model complete with transmission and logistics), there are several recommendations implied by the mathematical and solution structures developed herein and by our experience. This section contains specific recommendations regarding implementation of the transportation and logistics models worldwide. (These recommendations apply equally well to the interconnected transportation, supply, and demand models, i.e., the entire economic model).

1. Every last line of computer code must be written in a widely commercial, common, understandable, maintainable, contractible high-level language. We recommend that every single line of executable code in EIA's model be written in C# or C++ or similar under a widely commercially available compiler. No other type of code is maintainable or broadly accessible and interpretable and should be precluded. We recommend Microsoft Visual Studio C# because it is so widely used and so easy to use. EIA should avoid any and all code written in Visual Basic (VB), MatLab, GAMS, AMPL, FORTRAN, Pascal, or any other noncommercial or arcane language. C#, C++, Java, Python, etc., are the industry standard software engineering languages. Every last line of executable code should be written in a native high-level language like these. We especially recommend against VB code. VB code is notorious for being private, undocumented, undocumentable, and highly specialized to its developer. (VB codewriters do whatever they want in the privacy of their own offices.) We dub VB "selfie" code, private do-it-yourself code that simply is not universally accessible, maintainable, or understandable. EIA cannot afford to rely on obsolete programming languages that cannot be easily purchased on the open market and widely supported. (Did we hear that Microsoft is considering retiring VB for this reason and replacing it with C#?)

2. EIA should avoid basing the architecture on "bundling equations and data for delivery to a third party solver." We recommend EIA avoid solvers altogether, and this is consistent with our recommendation that EIA avoid complementarity and linear programming altogether for your transportation and logistical needs. This report tells you why they are unnecessary; you can analytically rather than numerically solve complementarity equations and develop the network microeconomic equilibrium equations right from the outset. You never need those ponderous and bulky solution methods. We have shown herein that the implementation is far larger than necessary in the cases of complementarity and linear programming, and eminently cumbersome. The words "big" and "cumbersome" are watchwords for high error rate, extreme labor intensity, and slow cycle time. Bundling code and data for a nontransparent solver is undesirable in modern computer engineering practice. It is literally impossible to write a user-oriented system that bundles code and data to deliver to a concentrated application like a solver. If EIA does that, your modeling software will remain highly labor intensive, specialized, and opaque, and it is unlikely to be broadly used. Furthermore, EIA cannot look inside the solver, meaning EIA can never be sure what the central calculations of its models are truly doing. Even if EIA could look inside a solver, the algorithms and programming are often so Spartan that they are impossible to understand, diagnose, and maintain. Have you ever looked inside the source code of say a matrix inversion program, a linear program, or a line search? The programming alone is daunting.

3. This report shows that linear programming and complementarity are unnecessary; they simplify to network microeconomic equilibrium when you perform the analytic substitution. Network microeconomic equilibrium has such simple and direct solution algorithms that you can implement them independently for every node in your model and can observe their operation continuously from iteration to iteration. There is no need with such an algorithm to bundle code and data to deliver to a "solver." The network microeconomic solver is simple and visible and programmed in C#. And once you solve the network microeconomic problem, you have de facto solved the complementarity problem. Network microeconomic equilibrium is the best way to solve the complementarity equations.

4. EIA should absolutely maximize the use and reliance on **analytic solutions to optimization problems** so that EIA absolutely minimizes the amount of computer programming, which is known to be intense and risky. The pace and accuracy of ongoing, real-time programming or equation specification for model building is simply too slow. The accuracy is too low. Analytics are much, much faster, non-redundant, and intrinsically insightful along with being maximally efficient from a numerical viewpoint. EIA should maximize the use of analytics in its nonlinear solutions so that it can control and minimize computation.

5. Disaggregation of your transportation and logistical assets in the market is at a high premium. Complementarity and linear programming are oversized and difficult in the way they represent network oriented, distributed transportation and logistical systems. This report has made that clear and shown why. That has been a major theme and mathematical proof in this report, the notion of the Cartesian product of hubs plus links times hubs plus links, an absolutely monstrous number for the size and scope of transportation problems EIA needs to solve. There is no escape from the inefficient, opaque techniques required to implement network problems within complementarity or linear programming. By eliminating complementarity and linear programming from consideration, EIA is thereby not subjecting yourself to the inflexibility and ponderousness of the implementation itself limiting your ability or propensity to disaggregate. Limitations on the size and ease of use have already occurred with linear programming. (If it hadn't, EIA would long ago have had an integrated world gas model and an integrated world oil model, both with massive transportation matrices, functioning rapidly and smoothly. We had one in 1990.) It is simply too hard to create a reliable, large-but-sparse constraint matrix that actually solves a suitably large problem. The same impediments occur with complementarity, and to an even greater degree. This report shows why. Complementarity is "hubs plus links times hubs plus links" in size, and linear programming is "hubs times activities" in size. Complementarity and linear programming algorithms are "full rank," oversized algorithms for the need, as we have shown. That isn't a matter of conjecture; it is demonstrated mathematically herein. That aspect of those methodologies, shown mathematically herein, will assuredly limit EIA's model size (and thereby the allowable degree of disaggregation) and efficiency of operation. EIA has endured that for too long with linear programming, and it is time to augment your productivity.

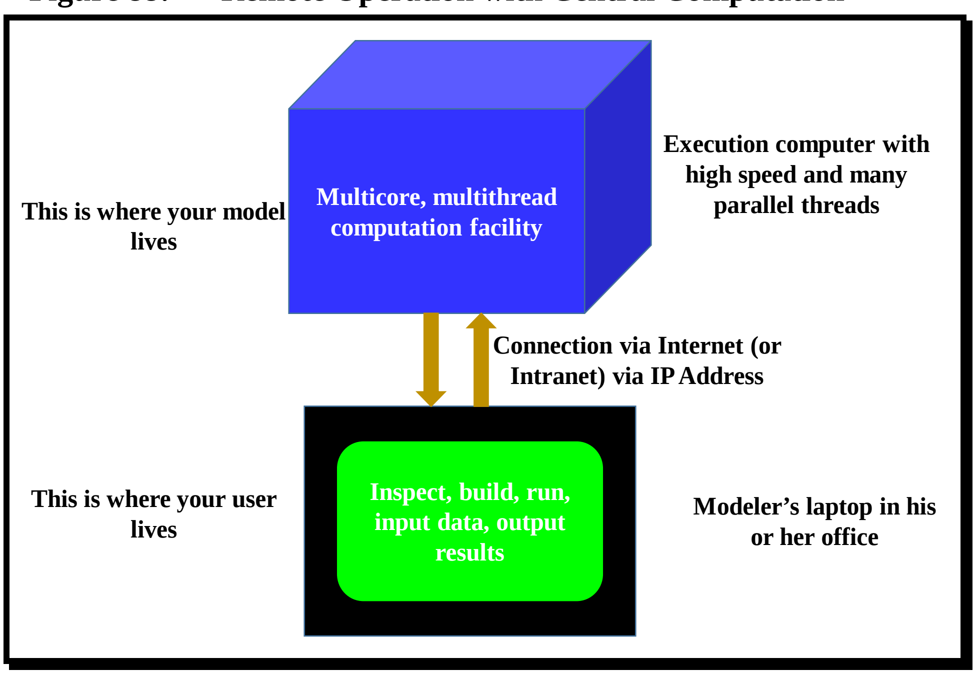
6. EIA models must be graphical and visual. They must be driven by a fully graphical representation. The picture of the model **is** the model. Managing your model by managing the pictures is an ideal way for EIA to augment model productivity. The computer science joke,

entirely correct, is "A picture is worth 1024 words." EIA needs to manage pictures, not computer programs or commands in specialized, arcane languages.

7. EIA models must be geographically hierarchical. The world gas market and the world oil and products market are geographically hierarchical. EIA's data are all naturally organized in a geographically hierarchical form. To illustrate, one of your regions is North America. Under North America are the United States, Canada, and Mexico. Under Canada are the various provinces (British Columbia, Alberta, Saskatchewan, Ontario, etc.) Under Alberta, we will see supply, demand, transportation, refining, and storage. Under demand, we will see core demand, electric generator demand, and other categories. You can already sense a natural hierarchy of data indexes emerging. EIA needs to represent this geography and segmentation hierarchically exactly as is done within a GIS. If you do so, you will get map display for free in your regions, and you can navigate regions with aplomb. Your country or regional data will be part of the overall hierarchy, which becomes a data management as well as model organizing hierarchy. That easily admits of modern database products and GIS products.

8. In the modern world of computation, you do not want to require the model to execute on the same machine that you manage the structure and data. (That is the standard modus operandi with C#, C++, and other systems. They default to running everything entirely on one computer.) You simply cannot have the Microsoft Office and other high-maintenance, personalized, distributed-user programs running on every large, parallel computer you have. You must have multi-user access to your computational computers via the Internet (or the Intranet if you insist on a closed internal network.) The configuration you want is that in Figure 88.

Figure 88: Remote Operation with Central Computation



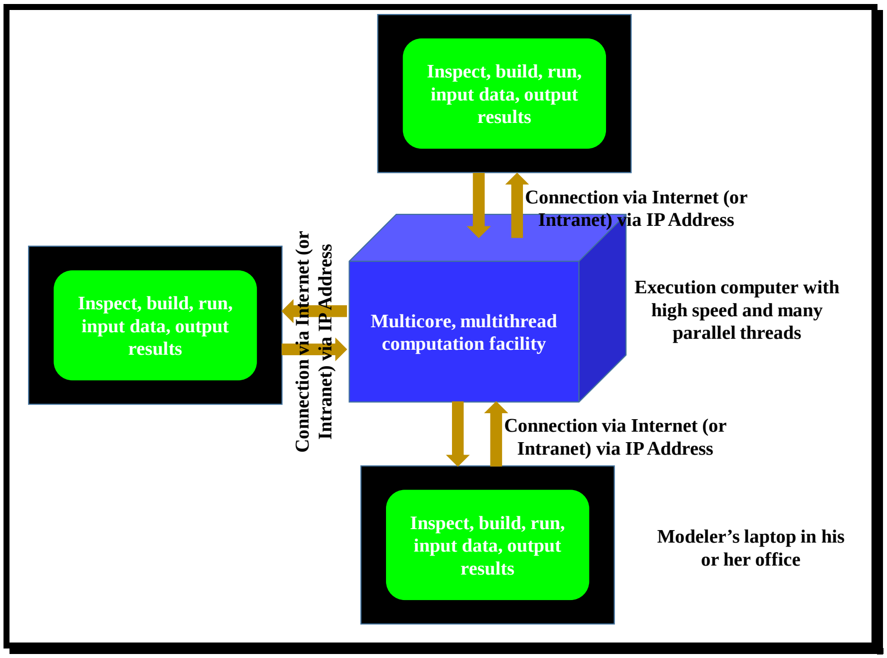
This configuration allows model network management, data input, data review, reporting, report retrieval, execution, simple modification, and gross modification to be done strictly in a browser and strictly on a computer remote from the computational computer. It could be 50 feet away, or it could be several thousand miles away. (Such remote access must not be simplistic access. It must be pure Internet IP access, access in which the remote browser and its activities are totally decoupled from the main computational machine.) The model per se will be housed and run on a central computer. This way, your IT costs and lead time are minimal, and you will not be captive to or hampered by computer hardware or software changes. How many times has a change in Windows or Office brought your individual computers and applications running on them to a screeching halt or created or unearthed bugs? It has happened to us repeatedly when we ran single machine systems. (Wait until Windows 8.2 comes out or Office 2015 comes out!) You want to be in a position that the browser-based editor and graphical interface are self-contained, remote from the model, and maintained separately and individually as distributed computation. And that is why you want them programmed in native source code—C#, C++, or Java. You want your compute engine to be a pristine execution environment not bogged down with memory-sapping, processor-sapping, non-internal "cleanup," high maintenance, cluttered user programs. You will want it professionally maintained as a parallel computation site. You will want it to be on the "cloud" or to be an internal cloud emulator. (EIA and everyone else are likely to be on the cloud within half a decade. The cost will be so low that an internal, smaller scale system may be two orders of magnitude higher. The cost of parallel threads on the cloud is dropping like a stone.) You will want your professionals to be able to work remotely 24/7 from any location in the world. In the modern world, organizations have found out that if they offer this degree of access, people work more than a standard schedule by their own volition and the productivity of their labor skyrockets.

You will have to manage the security and access to your main model, but this is a long since solved, standard problem. The real reason you want to do this is illustrated in Figure 89—multiuser remote access via the Internet. There is no substitute for this in the modern world.

9. EIA needs to use implementation contractors who can do it **for** you and **with** you. There are organizations that can do this and have done this. ArrowHead has done this, and we have a system configured like this. You need vendors who can deliver or develop software and at the same time familiarize and teach EIA people to augment autonomy and independence. EIA does not want software vendors such as those who built the Obamacare website, vendors with promises rather than demonstrated success. EIA wants vendors who have demonstrably done this commercially.

10. The configuration in Figure 88 and Figure 89 is the only way EIA can systematically parallelize your calculation. You need to systematically parallelize your calculation so that you can expand the level of transportation, storage, and infrastructural detail literally ad infinitum and remain certain that you will always have or be able to access the compute power to solve it.

Figure 89: Multiple Remote Access via the Internet



11. This hardware-software configuration is an ideal way for EIA to share its models with third parties if EIA should decide to do so. You can simply open up an IP address to a third party, and—voila! They have access to your workproduct and model data set. EIA's models have in the past not been used much by third parties. They are too cumbersome. Access and interaction by third parties can pay benefits to EIA far more than you might anticipate. Third parties will work with you! They will make suggestions. They will review. That can only help EIA.

12. Do not allow the implementation, configuration, knowledge, or modeling specialization to be so tightly held or arcane within your organization that only a tiny cadre of inside people are the only ones who can access, inspect, run, or report. Make the implementation, knowledge, and access so broad that anyone and everyone within EIA can run it. That way, they can play with it (not necessarily the official copy) and be sure they know what they are doing and what the organization is doing. The model has to be so transparent and so simple to run that everyone does and can do it. That ensures maximal use and understanding and minimizes cycle time. EIA has probably been captive to the "specialized knowledge" or "insider knowledge" or "indispensable staff" problem, and that is not a good strategy in the high productivity, modern world. EIA needs a bevy of trained, cognizant people so as not to be mired in critical path or critical employee problems. The original design of our system (an older incarnation not owned by ArrowHead) by the organization that paid us to build it was this: *"The model must be graphical and self-documenting. If I, the President of the company, want to look at it, I can. I can see the network and browse results. If the most junior technical analyst wants to look at it, he or she can. We both access it in the same way through the same portal, and we can both understand it."* That original model was dubbed the "Doomed to Failure" model, and yet it was a triple off the wall in terms of success and profitability to the sponsoring organization. It succeeded because of universal access and broadbased work. The modeling software would allow the Administrator to access, survey, and browse a mode, its structure, its data, and its

results. It should allow the modelers within EIA to build, modify, run, and report it. It would allow third party access. EIA should pursue that type of implementation so as to maximize access. (Not to worry; it is simple to secure and manage the official organizational copy of the model and data.)

13. EIA's implementation should **stay completely away from monolithic global welfare maximization, complementarity, and linear programming**. They are too dangerous, too arbitrary, too oversized, too error prone, and far too labor intensive. With such implementations, we have demonstrated that constraints can ruin the interpretation of shadow price as price. Insertion of welfare function modifications are entirely too easy, and they ruin the entire solution. EIA wants zero probability of either of these pernicious things occurring. A monolithic global welfare maximization approach will not serve EIA's needs. Monolithic global welfare maximization as an elemental concept is not appealing, and theorems and results and sufficiency conditions that try to equate it with direct approaches are not appealing. The notion of a central utility or welfare function, even the Samuelsonian welfare function, is alien to the way the real world works. Agents acting in their own interest is appealing; that is the way the real world works.

14. EIA absolutely must separate programming from real-time modeling. **You must not have a model in which you have to do any programming to implement, modify, execute, or report it.** You should not allow or require one single iota of programming, not one single line of code, for model building or modification. All programs must be pre-written before you get to the task of modeling. Network microeconomic modeling facilitates that goal. Complementarity conflicts with that goal.

    As an example, EIA should never allow yourself to have to write complementarity equations in order to build or modify your model. That alone should eliminate complementarity from all consideration. You cannot afford to write and organize equations to represent elements in your model or write and organize data commands for input to your model (the way GAMS and AMPL work). All coding must be pre-done and pre-completed within predefined, accessible objects. Your data must enter those objects from and to Excel or a database program such as Access or Oracle. We have heard the complementarians say: "We will write some equations to represent Persian Gulf suppliers." "We will write some equations to represent NW Europe/Rotterdam refiners." "We will write some equations to represent Padd 5 product demand." The words "write some equations" are literally profane. EIA must not let itself get into writing equations to represent elements of or players within your model. That will condemn EIA's model cycle time to remain in the multi-months range. EIA needs your model cycle time to be in in the hour range.

    So where should the programming occur? It occurs inside pre-programmed node objects. Those programs are written, vetted, debugged, tested, and approved in advance as legitimate node objects well in advance of any modeling. EIA writes the supply curves on output links and the demand curves on input links in C# and embeds them in the node objects. With network orientation, you simply drag and drop those "objects" into an interconnected model, which when fully interlinked comprises a system. The secret sauce of network modeling is that dragging and dropping IS writing equations. The equations are pre-programmed. Those equations are fully visible to EIA and in fact modifiable by EIA. To wit, EIA should have access to the node logic for inspection and modification, and EIA must have the ability to write new node objects. That capability is easy to provide.

As a corollary, EIA should avoid organizing an arcane, sparse, literally non-debuggable constraint matrix in order to build your model. You have had more than enough of that in the last two decades, and it has assuredly extended your model cycle time and labor productivity.

15. EIA should use licensed, supported, commercial code wherever possible. **<u>With regard to the "build-buy" decision, EIA should "buy."</u>** EIA does not seem to have chosen this route in the past save for linear programming "solvers" and Microsoft Office, and that has been a mistake. Self-built, self-designed, private code is a mistake. The leverage EIA gets on commercial, commercially supported, actively supported, tested, non-"selfie" software can be immense. Use of such code broadly outside as well as inside EIA can be a boon. EIA should pursue that option rather than internally developing custom, "one-off" code using a combination of contractors and in-house people. If EIA buys software, they can take delivery on Monday, be trained on Tuesday, and be building your network model and inputting your data starting on Wednesday. EIA can begin with an existing network model and amend it so that it is truly EIA's model floor to ceiling. That whole process of getting a meaningful model running would be four weeks or less. The amendments and improvements would occur in an ongoing fashion after that.

    Keep in mind, the code used to make EIA's calculations is not what EIA's model is. EIA's model is the **<u>network, the input data, the market structure, the results, the insights, the parameter settings,</u>** etc. It is not the code, the software. The software is merely a tool to compute a solution. The software is not the model. To use an analogy, EIA would not dream of building your own in-house version of Excel just so you could be independent of Microsoft. (EIA wouldn't want to be beholden to Microsoft, would you? You sure wouldn't want to share credit with them! How do you know there aren't bugs? How do you know they will be there to support you?) To sidestep the issue, EIA simply chooses not to define Excel as part of its model. Quite the contrary, EIA benefits from the massive, billions of dollars, ongoing investment and improvement by Microsoft in its Excel product and use by myriad unseen customers around the world. EIA chooses to define Excel not as part of its model but as a tool to help with its model. By that same analogy, EIA would benefit from licensing or rebuilding proven, commercial software to do the browser-based graphical and operational front end and proven, commercial software to do the operational, executional portions of the model. EIA can probably negotiate source code licenses with legacy rights for the critical pieces of capability it believes it needs so that non-support risks are eliminated.

    EIA's license should make provisions for contractor as well as employee access. Most licenses allow that. That is necessary for EIA to achieve inter-contractor competition in order to discipline the price it has to pay for various contractor services and to free EIA from reliance on a small few contractors. To wit, that frees EIA from being captive to one or a small handful of contractors who have insider knowledge of EIA's software and extract rents and thwart competition.

EIA needs to have the ability to program supply and demand curves it wants to implement and to disclose calculations and results to the public domain without restriction. That is extremely easy to secure as part of license. The license must allow EIA to write its own node logic and link it into the system. That is an easy capability to secure, this “plug and play” capability.

As a premier licensee, EIA will have a lot to say about the evolution and functionality of commercial code. EIA is not an immaterial licensee or customer.

16. EIA should eschew the “do it yourself” approach to software and work with operational, demonstrable software to build, house, and run your transportation, logistics, and in fact other models.

18 SUMMARY OF CONCLUSIONS

There is an old tune "Anything You Can Do," composed by Irving Berlin for the 1946 Broadway musical **Annie Get Your Gun**. The song is a spirited duet, with one male singer [sharpshooter Frank Butler] and one female singer [Annie Oakley] attempting to outdo each other in increasingly complex tasks.

[Annie:] Anything you can do, I can do better!
I can do anything better than you!
[Frank:] No you can't!
[Annie:] Yes, I can!
[Frank:] No, you can't!
[Annie:] Yes, I can!
[Frank:] No, you can't!
[Annie:] Yes, I can. Yes, I can. Yes, I can!

You can see it performed at http://www.youtube.com/watch?v=_UB1YAsPD6U. The purpose of this paper is to lift the discussion out of "Broadway musical" mode and into the scientific and mathematical realm.

Economists have been hearing this refrain for five decades from linear programming and complementarity people. "Yeah, but complementarity can do that too!" "Don't leave us out; we can do that too." "Yes I can, yes I can, yes I can." Linear programming and complementarity have generally played defense, not mastering groundbreaking problems but rather arguing for equal standing. Complementarians assert they can solve all the same problems that network microeconomic methods can proactively solve, generally without proof or elucidation. (The proof and elucidation are contained herein.) As we have shown herein, complementarity solutions are laden with massive, oversized, complex solution algorithms (nonlinear programming, linear programming, complementarity algorithms such as those referenced by Gabriel op. cit.), meaning modelers have markedly less time for problem structuring and formulation. We advise complementarians to group their equations and solve them analytically. When they do so, the analytical solutions they get will be the network microeconomic equilibrium equations! If they proceed analytically, they end up with network microeconomic equations! Complementarity equations are much better solved as network microeconomic problems.

What does it matter if monolithic global welfare maximization, written in the form of complementarity equations, can take a problem specified in English, translate the problem into Chinese, solve the problem in Chinese, and translate the answer back into English? This paper shows how monolithic global welfare maximization has taken a network microeconomic problem expressed in English (agent-based decentralized decision making), translated it into a monolithic global welfare maximization problem (Chinese), solved it in Chinese (complementarity or linear programming), and then translated the answer back to English (shadow prices and quantities). Why not just solve the transportation problem directly in English (network microeconomics)! It is easier, smaller, and more direct. We have shown that the direct, graphical, network-oriented, microeconomic approach is so much faster, smaller, and more direct. It can handle perhaps two orders of magnitude more detailed transportation problems, and with far less possibility of error and probably two orders of magnitude less effort, time, and cost. And in cases where complementarity is an accurate representation of the competitive equilibrium problem, it gets the same answer. All this occurs without the risk of a monolithic global welfare function or any constraints that can distort the meaning of prices or tempt modelers to dishonest model formulations or solutions driven by social welfare considerations.

REFERENCES

1. Andreu Mas-Colell, Michael D. Whinston and Jerry R. Green, **Microeconomic Theory**, Oxford University Press, Jun 15, 1995.

2. Brock, H.W., and Nesbitt, D. M., **Large Scale Energy Models—A Methodological Analysis**, Stanford Research Institute report to the National Science Foundation, 1977, copy available from the author.

3. Gabriel, Steven A., Conejo, Antonio J., Fuller, J. David, Hobbs, Benjamin F., and Ruiz, Carlos, **Complementarity Modeling in Energy Markets**, International Series in Operations Research and Management Science, Springer, 2013.

4. Geoffrey A. Jehle and Philip J. Reny, **Advanced Microeconomic Theory**, (Third Edition), Prentice Hall, Apr 30, 2011.

5. Hal R. Varian, **Microeconomic Analysis**, Third Edition, Norton, Mar 17, 1992.

6. Luenberger, David G., **Linear and Nonlinear Programming**, Second Edition, Addison-Wesley, May 1989.

7. Nesbitt, Dale M. and Scotcher, Jill N., "Spatial Price and Quantity Relationships in World and Continental Commodity Markets," **The Energy Journal**, 2009 Special Issue.

8. Samuelson, Paul A. (1952). "Spatial Price Equilibrium and Linear Programming," **The American Economic Review**, Vol. 42, No. 3 (Jun.), pp. 283-303.

9. Busch, Lauren K., **Draft Review of Natural Gas Models**, Leidos report, August 14, 2014.

10. Nesbitt, Dale M., "The Economic Foundation of Generalized Equilibrium Modeling," **Operations Research**, Vol. 32.1984, 6, p. 1240-1267

11. Jeffrey C. Williams and Brian D. Wright, **Storage and Commodity Markets**, Cambridge University Press, 1991

12. Friedman, James W., **Oligopoly and the Theory of Games**, Elsevier/North-Holland (1977).